**University of Alberta**

# Progress towards genetically encoded calcium sensors based on circularly permuted monomeric red fluorescent proteins

by

Yankun Li

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Chemistry

Edmonton, Alberta
Fall 2007

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

*Your file Votre référence*
*ISBN: 978-0-494-33298-6*
*Our file Notre référence*
*ISBN: 978-0-494-33298-6*

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

*This thesis is dedicated to my lovely husband, Chengjie, and my daughter, Kaylee, for their unwavering support, understanding, strength and love.*

*Also to my parents for their constant encouragement and for instilling me with a belief in my own abilities*

# ABSTRACT

Circularly permuted protein variants are proteins in which the normal N- and C- termini have been connected with linker peptides of appropriate length and have been introduced new termini into surface-exposed loop regions. Some, but not all, circularly permuted proteins can still fold to a biologically active conformation after introduction of the new termini. It has previously been reported that circularly permuted variants of the green fluorescent protein (GFP) from *Aequorea Victoria* are able of chromophore formation and thus retain their intrinsic fluorescenc. Here we demonstrate that the red fluorescent protein know as mCherry can retain its fluorescence despite a variety of circular permutations at two different sites or insertion of 15 bp at six different sites which are located at a solvent-exposed loop between β-strands. The tolerance of mCherry for circular permutations and insertions provide a foundation for creating new genetically encoded indicators for important biochemical and physiological signals.

## ACKNOWLEDGMENTS

First of all, I would like to thank my supervisor, Dr. Robert E. Campbell, most sincerely for granting me the opportunity to work with him, and seeing me through to the submission of my master degree. I have appreciated his invaluable guidance, inspiration, advice, friendship and tolerance, during not only my graduate studies but throughout our working relationship. I have also appreciated his guidance and many informal 'thesis discussions'. His careful critique and enthusiastic encouragement to press on towards the goal was truly appreciated.

I also would like to thank sincerely the other members of my supervisory committee: Professor Liang Li and Professor Gordon Chan from Department of Oncology, for their active participation during my oral examination and for their thorough reviews and comments on this thesis.

On a personal note, I would like to thank my many friends and colleagues in the University of Alberta help in large or small ways has made my graduate study smoother and more wonderful. To all the members of Dr. Campbell's group, past and present, many thanks for being a great bunch to work with and for their helpful discussions; without their calming influence, good humor and friendship I could have finished the thesis a lot earlier, but been the worse for it! Especially to Huiwang and Zihao thanks for their years of friendship, encouragement and endless cliches to help me realise that I was actually making progress. I also thank Mr. Michael Carpenter for his thorough proof-reading of this thesis. With his help I eliminated many questionable grammatical endeavors

into the English language.

I would like to thank the personnel in the general, purchasing, and post offices, and the electronics and machine shops at the Department of Chemistry of the University of Alberta. Without their kind help, this research work would not have been possible.

I would like to thank the Department of Chemistry and the University of Alberta for providing me with the opportunity to learn and grow as a scientist, and for providing a safe and well-equipped facility to conduct my experiments.

Finally, I especially thank my husband, Chengjie, for his love, support, patience and encouragement over the course of my graduate studies, my daughter, Kaylee, for the happiness she has brought to me since she was born about one year ago, and my parents for their grateful help in taking care of Kaylee.

# TABLE OF CONTENTS

**ABSTRACT** .....

**ACKNOWLEDGMENTS** .....

**TABLE OF CONTENTS** .....

**LIST OF FIGURES** .....

**LIST OF TABLES** .....

**ABBREVIATIONS AND SYMBOLS** .....

**Chapter 1 Introduction: Fluorescent proteins as labels and biosensors** ..... - 1 -

1.1 *Aequorea* Green Fluorescent Proteins (GFP) ..... - 6 -

1.1.1 Discovery of *Aequorea* GFP ..... - 6 -

1.1.2 Structure of *Aequorea* GFP ..... - 7 -

1.1.2.1 Primary Sequence from Cloning ..... - 8 -

1.1.2.2 The chromophore structure of GFP ..... - 8 -

1.1.2.3 Crystal Structures ..... - 8 -

1.2 *Anthozoan* Fluorescent Proteins ..... - 12 -

1.2.1 Discovery of Fluorescent Proteins in Corals ..... - 12 -

1.2.2 Properties of DsRed ..... - 13 -

1.2.2.1 Spectroscopic properties ..... - 14 -

1.2.2.2 Oligomerizaion ..... - 14 -

1.2.2.3 Slow maturation ..... - 15 -

1.2.3 Structure of DsRed ..... - 16 -

1.2.3.1 Primary Sequence from Cloning ..... - 16 -

1.2.3.2 Chromophore Structures in *Anthozoa* GFP Homologs ..... - 16 -

1.2.3.3 The crystal structure of *Anthozoa* proteins ..... - 19 -

1.3 Application in Fluorescence microscopy ..... - 19 -

1.3.1 Introduction to fluorescence microscopy ..... - 20 -

1.3.1.1 Fluorescence ..... - 20 -

1.3.1.2 Fluorescence Microscopy ..... - 23 -

1.3.2 Fluorescent proteins vs. bioluminescent proteins vs. small molecule fluorophores for calcium sensing in live cells ..... - 24 -

1.4 Engineering of Fluorescent Proteins ..... - 30 -

1.4.1 Engineering New Colors ..... - 30 -

1.4.1.1 *Aequorea* GFP Variants ..... - 30 -

1.4.1.2 *Anthozoan* Proteins ..... - 34 -

1.5 Engineering "Structural" Variants ..... - 36 -

1.5.1 Fusions ..... - 36 -

1.5.2 Insertions ..... - 36 -

1.5.3 Circular Permutations ..... - 37 -

1.6 Engineering Sensors ..... - 40 -

1.6.1 FRET Sensors ..... - 40 -

1.6.2 Single FP Sensors ..... - 42 -

1.6.2.1 Camgaroo indicators ..... - 42 -

1.6.2.2 Pericams ..... - 43 -

1.6.3 Translocation Sensors ..... - 44 -

1.7 Objectives of the thesis ..... - 44 -

1.8 Literature Cited ..... 45

**Chapter 2 Selection of mCherry Variants with a Peptide Insertion ..... 54**

2.1 Introduction ..... 55

2.2 Experimental ..... 60

2.2.1 General methods and materials ..... 60

2.2.2 Gene Construction ..... 60

2.2.3 Random 15 bp Insertion Mutation Library of mCherry Construction ..... 61

2.2.4 Library Screening ..... 63

2.2.5 15 bp Insertion Mapping by Colony PCR or Digest ..... 63

2.2.6 Protein purification and characterization ..... 64

2.2.7 Spectroscopy of the Insertion Protein and the Fluorescence Intensities of the Colonies ..... 65

2.3 Results and Discussion ..... 66

2.3.1 Construction of mCherry in pUC18 Plasmid ..... 66

2.3.2 Construction of 15 bp insertion mCherrry Variants ..... 67

2.3.3 Characterization of Insertion Variants of mCherry ..... 83

2.4 Conclusions ..... 87

2.5 Literature Cited: ..... 88

**Chapter 3 Circularly Permuted Variants of mCherry and Progress towards Genetically Encoded Calcium Sensors ..... 90**

3.1 Introduction ..... 91

3.2 Experimental ..... 95

3.2.1 General methods and materials ..... 95

3.2.2 Gene Construction ..... 95

3.2.2.1 Construction of Circularly Permuted mCherry ..... 95

3.2.2.2 Construction of Camgaroo-type Gene....96
3.2.2.3 Construction of Pericam Gene....100
3.2.3 Bacterial Transformation and Protein Purification....100
3.2.4 Library Construction, Mutagenesis and Library Screening....101
3.2.5 Spectroscopy of the Circularly Permuted mCherry Variants....102
3.3 Results and Discussion....102
3.3.1 Construction of Circularly Permuted Variants of mCherry....102
3.3.2 Spectroscopic properties of the circularly permuted variants....103
3.4 Conclusions....111
3,5 Literature Cited....114
**CHAPTER 4: Conclusions and Future Work....117**
**APPENDIX A: Tubes and labels....122**

# LIST OF TABLES

**Table 1.1** Comparison of the characteristics of the fluorescent probes: small molecules, aequoin and fluorescent proteins. The more +, the more advantages. ..... - 29 -

**Table 2.1** The characteristics of mCherry insertion variants compared to those of mCherry ..... 86

**Table 3.1** The primers for the second PCR to construct the gene of circularly permuted mCherry (cpmCherry). The underlined letters show the duplicated gene at both the N- and C- temini of cpmCherry. The bold letters encode the stop codon gene and the italic letters are the restriction sites (In 5' primer is Xba1 and in 3' primer is EcoR1.) ...97

**Table A.1** cpmCherry variants gene and the gene would be used to make mCherry $Ca^{2+}$ sensors..... 123

**Table A.2** mCherry insertion variants gene and the gene was used to make the insertion variants. ..... 124

**Table A.3** Primers were used to in this thesis working. ..... 126

# LIST OF FIGURES

**Figure 1.1** FP-based biosensor design. A. FRET-type biosensor. B. Single FP-type biosensor with an inserted conformationally responsive domain within FP. C. Single FP-type biosensor with an inserted conformationally responsive domain within cpFP. D. Single FP-type biosensor with a fused conformationally responsive domain at the new N- and C- termini of cpFP. ..... - 6 -

**Figure 1.2** Mid-summer specimens of *Aequorea Jellyfish* photographed in natural environment (A, B & C) and in seawater supplemented with KCl in a darkroom (D) at the University of Washington's Friday Harbor Laboratories [23]. ..... - 7 -

**Figure 1.3** GFP sequences. (Lines 1 and 2) WT GFP and EGFP gene, respectively, Clontech Laboratories. (Line 3 and 4) WT GFP and EGFP amino acid sequence, respectively. (Lines 5) Numbering of amino acids of both EGFP and WT GFP. The inserted Val in EGFP is numbered 1a to maintain correspondence with the WT GFP numbering. ..... - 10 -

**Figure 1.4** Mechanism proposed by Cubitt et al. [29] for the intramolecular biosynthesis of the GFP chromophore. ..... - 10 -

**Figure 1.5** A ribbon representation of the three-dimensional structure of GFP[33], showing 11 β-strands forming a hollow cylinder through which is threaded a helix bearing the chromophore, shown in Van der Waals Radii representation. ..... - 11 -

**Figure 1.6** Amino acid side chains, main chain carbonyls and amides, and solvent waters in the immediate vicinity of the chromophore of S65T GFP [33]. Side chains are labeled with the one-letter code for the amino acid and the residue number. Main chain

groups are labeled with the residue number. Water oxygens are denoted by W and the corresponding residue number in the PDB file. Hydrogen bonds are shown as dotted lines labeled with the distance between the heteroatoms in angstroms. From reference [8]. ..... - 12 -

**Figure 1.7** Sequences. (Lines 1 and 2) WT DsRed and mCherry gene, respectively, Clontech Laboratories. (Line 3 and 4) WT DsRed and mCherry amino acid sequence, respectively. (Lines 5) Numbering of amino acids of both WT DsRed and mCherry. - 18 -

**Figure 1.8** Scheme for the formation of mature red chromophore of DsRed by conversion from the green intermediate by oxidation. Proposed by Gross et al. [62]. - 18 -

**Figure 1.9** Stereoview of the three-dimensional structure of DsRed showing 11 β-strands forming a hollow cylinder through which is threaded a helix bearing the chromophore; shown in Van der Waals Radii representation. ..... - 20 -

**Figure 1.10** Jablonski diagram. ..... - 22 -

**Figure 1.11** Diagrams of transmitted light microscopy (A) and epi-illumination fluorescence microscopy (B). ..... - 25 -

**Figure 1.12** The structure of fura-2, pentapotassium salt (A) and Fura Red, tetrapotassium salt (B). ..... - 27 -

**Figure 1.13** A. Genetic construction of a chimera. B. Illustrative example in which the fusion protein targets to the nuclear membrane. ..... - 29 -

**Figure 1.14** The chromophore of WT *Aequorea* GFP (A) and EGFP (B). Abbreviations: $\lambda_{ex}$, excitation wavelength in nm and $\lambda_{em}$ emission wavelength, which will be used in Figures 1.15 to 1.19. ..... - 31 -

**Figure 1.15** The chromophore of the original BFP. ..... - 32 -

**Figure 1.16** The chromophore of CFP or Cerulean ..... - 33 -

**Figure 1.18** The chromophore of DsRed. ..... - 34 -

**Figure 1.19** The chromophore of mRFP1. ..... - 35 -

**Figure 1.20** The principle of circular permutation of a polypeptide chain. The N- and C- termini of the wild-type protein are connected, generating a circular protein intermediate. The protein backbone is then cleaved at a different position, yielding a circularly permuted variant of the protein with novel N- and C- termini. ..... - 39 -

**Figure 1.21** The general idea of engineering circularly permuted mCherry created from insertion variants. ..... - 40 -

**Figure 1.22** Jablonski diagram to show the principle of (A) fluorescence (B) FRET.- 41 -

**Figure 2.1** Flow chart of the MGS protocol of *in vitro* transposition reaction.....58

**Figure 2.2** Structure of the 15 bp insertion generated by the MGS. (A) Target DNA; (B) Target DNA with Entranxeposon insertion (transposition complex); (C) Target DNA with 15 bp insertion. .....59

**Figure 2.3** DNA agarose gel. All samples were subjected to electrophoresis in a 1% agarose gel in a buffer containing 100 mM Tris-HCl, 120 mM sodium borate (pH 8.4), 0.1 mM EDTA, and ethidium bromide (0.5 mg/ml). Line L: 1 Kb plus ladder (Invitrogen). Line 1: pUC18 contains CFP. Line 2: pUC 18 vector which is from the digested fragment of pUC18 containing CFP digested with Xba1 and EcoR1. Line 3: pBAD contains mCherry. Line 4: mCherry gene with Xba1, Kpn1 and EcoR2 restriction site. Line 5: the ligation product of ligating pUC 18 (Line 2) and (Line 4). Line 5: digest of the ligation product with XbaI and EcoRI. Line 6: digest of the ligation product with KpnI. ..... 67

**Figure 2.4** Pool of target mCherry clones carrying an Entranceposon insertion at random locations. is mCherry gene, is pUC18 vector. A, B and C, insertions occur within mCherry gene. D and H, insertions occur between mCherry and pUC18 vector. E, F and G, insertions occur within pUC18 vector. ..... 69

**Figure 2.5** Gel electrophoresis patterns of Entranceposon insertion mCherry digest. All samples were subjected to electrophoresis in a 1% agarose gel in a buffer containing 100 mM Tris-HCl, 120 mM sodium borate (pH 8.4), 0.1 mM EDTA, and ethidium bromide (0.5 mg/mL). Lane 1 of each gel: 1 kb plus DNA ladder (Invitrogen in A and Fermentas in B). A shows the pattern for the fragments of and Entranceposon insertion mCherry variants digested by EcoR1 and Xba1. B shows the fragments of Entranceposon insertion mCherry digested by Not1. is mCherry gene, is pUC18 vector, is the isertion gene, Entranceposon. Shows the band of interest which was cut from the gel. ..... 70

**Figure 2.6** Cartoons of 1) Not1 digest cuts out the body of Entranceposon; 2) 15 bp insertion in the mCherry plasmid; and 3) Digest mapping of the 15 bp insertion in mCherry plasmid. ..... 71

**Figure 2.7** PCR mapping of the 15 bp insertion in mCherry plasmid.....71

**Figure 2.8** Gel electrophoresis patterns of PCR products or digest fragments with mapping of the 15 bp insertion mCherry. All samples were subjected to electrophoresis in a 1% agarose gel in a buffer containing 100 mM Tris-HCl, 120 mM sodium borate (pH 8.4), 0.1 mM EDTA, and ethidium bromide (0.5 mg/mL). A-C are mapping by PCR. A and B, three PCR (a, c and b (Figure 2.7) in order) were performed. C, two PCR (a and b) were performed. D-M are mapping by digest. D and F, two digests (a and b (Figure 2.6) in order) were performed. Others, three digests (c, a and b in order) were performed. G-M, the first three columns were the #01, which were confirmed by sequence, as reference. .....74

**Figure 2.9** The sequence of the insertion mCherry variants compared to that of mCherry. 15 bp, translating to 5 amino acids, (gray) were inserted within mCherry, among which 5 bp (italics and bold) were duplicated from the gene of mCherry and the other 10 bp (tgcggccgca) remained from the Entranceposon, containing gcggccgc Not I restriction site. A, is #1 in which AAAQK was inserted between K138 and K139 within mCherry. B, is #2 in which MRPQQ was inserted between Q137 and K138 within mCherry. C, is #19 in which DAAAH was inserted between H25 and E26 within mCherry. D, is #32 in which AAASV was inserted between V22 and N23 within mCherry. E, is #47 in which CGRTY was inserted between Y193 and N194 within mCherry. F, is #67 in which NAAAA was inserted between A184 and K185 within mCherry. .....81

**Figure 2.10** Locations of 15 bp insertions in mCherry are shown in a schematic drawing of the overall fold of mCherry. Circles "O" are the insertion locations, the open star is the chromophore, residues 66-68, which is in the central helical segment (residues 58-72). .82

**Figure 2.11** Locations of 15 bp insertions in mCherry are shown in a cartoon drawing of mCherry. Balls are the insertion locations. A, B, C, D, E and F are #1 (138-139), #2 (137-138), #19 (25-26), #32 (22-23), #47 (193-194) and #67 (184-185) insertion mCherry variants. ..... **82**

**Figure 2.12** The relative fluorescence spectral of purified insertion mCherry variants and mCherry. All the measurement were made under the same conditions and all the proteins had the same concentrations. ..... **85**

**Figure 3.1** Scheme for rational and random generation of circularly permuted variants of a protein. Both approaches utilize a circular gene that is constructed by cleavage at Kpn1 restriction sites flanking the natural gene (a) and subsequent ligation with T4 DNA-ligase (b). For random circular permutation, the circular gene is partially digested with DNase I (c) and a library of relinearized genes is isolated. The library is treated with T4 DNA ligase and T4 DNA polymerase (d), yielding blunt-ended, repaired fragments that can be cloned into an appropriate expression plasmid. The circular gene can also be used as a template for PCR amplification to generate rationally designed circularly permuted variants (e, f). ..... **92**

**Figure 3.2** (A) shows the sequence insertion in mCherry gene. 10 bp (bold) remnant of the Entranceposon, containing gcggccgc NotI restriction site. 5 bp (italics) were duplicated from the gene of mCherry. ^ is the digestion site. (B) is the gene of CaM/M13 after digestion with NotI. (C) is the mCherry vector of No. 1, 19, 32 and 67 after digestion with NotI without the additional bases. (D) is the mCherry vector of No. 2 after digestion with NotI. (E) is the mCherry vector of No. 47 after digestion with NotI. ..... **99**

**Figure 3.3** Locations of new termini in fluorescent, circularly permuted mCherry variants are shown in: a schematic drawing of the overall fold of mCherry modified to show the starting points of fluorescent circular permutations (O), the linker (GGTGS) connecting the original N- and C- termini, and the approximate location of the chromophore (open star residues 66-68), which is in the central helical segment (residues 58-72).....**104**

**Figure 3.4** (A)The emission spectra of 183cpmCherry and 192cpmCherry, which were expressed at 37 °C. The excitation was 480 nm (labeled Ex @ 480) and 540 (labeled Ex @ 480) separately. The spectra were normalized based on the green peak of the spectra excited at 480nm. (B) The emission spectra of mCherry.....**106**

**Figure 3.5** The emission spectra of 192cpmCherry. 192-18 is 192cpmCherry expressed at 18 °C; 192-37 is 192cpmCherry expressed at 37 °C; and 192-1.2-37 is 192cpmCherry1.2 expressed at 37 °C. The spectra were normalized based on the green peak of the spectra excited at 480nm. .....**107**

**Figure 3.6** The emission spectra of 183cpmCherry and 192cpmCherry. 183-18 is 183cpmCherry expressed at 18 °C; 183-37 is 183cpmCherry expressed at 37 °C; 183-M-37 is 183cpmCherry1.5 expressed at 37 °C; and 183-1.5-37 is the mutated 183 circularly permuted mCherry2 at 183 site expressed at 37 °C. The spectra were normalized based on the green peak of the spectra excited at 480 nm. .....**108**

**Figure 3.8** The *in vitro* emission spectra of 183cpmCherry variants expressed at 37 °C and stored at 4 °C for months. 183-1.5 is 183cpmCherry1.5 and 183-2 is the circularly permuted mCherry2 at 183 site. (A) and (B) were normalized based on the green peak of

the spectra excited at 480nm. (C) was normalized based on the red peak of the spectra excited at 480nm. .....110

**Figure 3.9** Topologies of FP, cpFP, and chimeras with other proteins. The other proteins are depicted schematically as moon and ellipse. (A) The original FP molecule is represented as a cartoon drawing; (B) A tandem fusion product of FP with another protein; (C) the insertion of FP into a split protein domain; (D) insertion of a complete protein domain into the β-barrel backbone of FP; (E) cpFP connected the original termini with a loop and produced new termini; (F) A tandem fusion product of cpFP with another protein; (G) the insertion of cpFP into a split protein domain; (H) insertion of a complete protein domain into the β-barrel backbone of cpFP.....113

## ABBREVIATIONS AND SYMBOLS

| | |
|---|---|
| BAP | Bacterial Alkaline Phosphatase |
| BFP | Blue Fluorescent Protein |
| bp | Base Pair |
| BPTI | Bovine Pancreatic Trypsin Inhibitor |
| CaM | Calmodulin |
| CFP | Cyan Fluorescent Protein |
| CP | Non-Fluorescent Chromoprotein |
| cp | Circularly Permuted |
| CR domain | Conformationally Respinsive domain |
| CyPet | Cyan Fluorescent Protein for Energy Transfer |
| DAPI | 4',6'-diamidino-2-phenylinde |
| DsRed | Discosoma Sp. Red |
| DAPI | 4',6'-diamidino-2-phenylinde |
| EBFP | Enhanced BFP |
| EC | Exitation Coefficient |
| ECFP | Enhanced CFP |
| EGFP | Enhanced GFP |
| ER | Endoplasmic Reticulum |
| EYFP | Enhanced YFP |

| | |
|---|---|
| E.coli | *Escherichia coli* |
| FITC | Fluorescin Isothiocyanate |
| FP | Fluorescent Protein |
| FRET | Fluorescence Resonance Energy Transfer |
| GFP | Green Fluorescent Protein |
| Kan | Kanamycin |
| Kb | Kilo Base pairs |
| LB | Luria-Bertani |
| MGS™ Kit | The Mutation Generation System™ Kit |
| mRFP | Monomeric RFP |
| PCR | Polymerase Chain Reaction |
| PMT | Photomutiplier Tube |
| QY | Quantum Yield |
| RFP | Red Fluorescent Protein |
| SHM | Somatic Hypermutation |
| td tomato | Tamdem Dimeric dTomato |
| UV | ultraviolet |
| WT | Wild Type |
| YC3.3 | Yellow Cameleon 3.3 |
| YFP | Yellow Fluorescent Protein |
| Ypet | YFP for Energy Transfer |
| ZFP | Zoan YFP from Zoanthus Sp. |

# CHAPTER 1

# INTRODUCTION:

# FLUORESCENT PROTEINS AS LABELS AND BIOSENSORS

In nature, living organisms exhibit a great variety of fluorescent and color appearances. Generally, the enormous variety of colors and fluorescent hues displayed by living organisms are determined by low molecular weight pigments or chromoproteins. Chromoproteins are proteins in which a small nonpeptide molecule or metal ion (a prosthetic group) binds to the proteins and is essential for the chromogenic properties of the proteins [1, 2]. Fluorescent proteins (FPs), which have existed in jellyfish and various species of coral for more than 160 million years [3], represent the only known exception to this rule; they can form internal chromophores without requiring accessory cofactors, external enzymatic catalysis or substrates other than molecular oxygen [4]. For example, aequorin need cofactor, coelenterazine.

Following the cloning of the green FP (GFP) from the jellyfish *Aequorea victoria* in 1992 [5], and the first demonstrations that it could be expressed in organisms other than jellyfish [6, 7], GFP rapidly established itself as an indispensable tool of modern cell biology. The primary reason for this popularity is that genetically inserting GFP into a living cell or tissue enables the use of fluorescence microscopy to "see" biological structures that would otherwise not be readily visualized. The usefulness of GFP has been further expanded through the engineering of blue (BFP), cyan (CFP), enhanced green (EGFP) and yellow (YFP) variants which provide the unprecedented ability to perform multicolor imaging of fusion proteins and fluorescence resonance energy transfer (FRET) measurements in living cells [8-12].

In 1999 it was reported that corals are an abundant source of FPs [12], among which, drFP583 from *Discosoma* coral (a.k.a. DsRed) has spectral characteristics dramatically red-shifted from GFP. Efforts to improve the tetrameric wild type (WT) DsRed resulted in a first generation monomeric red fluorescent protein known as mRFP1 [10]. Further development of mRFP1 led to the mFruit series of FPs [13, 14], of which mCherry1 and mCherry2 [14] are the variants most relevant to this thesis work.

The use of FPs as static fluorescent labels is a ubiquitous technique that has been employed in thousands of published reports. This is in stark contrast to the relatively few examples of FRET-type (Figure 1.1A) or single FP-type (Figures 1.1B, C and D) dynamic fluorescent biosensors. FRET-type biosensors consist of a CFP and a YFP fused to a domain that undergoes a structural change in response to a specific event, such as binding of a small molecule or ion (e.g. $Ca^{2+}$), or a post-translational modification (e.g. phosphorylation) [9, 15, 16]. The rearrangement of the conformationally responsive domain changes the FRET efficiency, and the ratio of cyan to yellow fluorescence is modulated.

Single FP-type biosensors consist of an FP with an inserted conformationally responsive (CR) domain, such as calmodulin (a $Ca^{2+}$ binding domain) [17-20]. The conformational change that occurs upon binding the analyte of interest (e.g. $Ca^{2+}$) can change the environment of the highly sensitive GFP fluorophore and thereby result in a change in the fluorescence intensity [17-20].

In this thesis work, transposon-mediated insertional mutagenesis was used to identify six locations within the β-barrel of mCherry which tolerate insertion of five amino acids without abolishing the fluorescence. For each insertion variant a circularly permuted (cp) version (a topological variant) was created with new N- and C-termini at the position of the insertion (and a linker between the original N- and C- termini). Based on the cpmCherry variants, development of a single FP red fluorescent $Ca^{2+}$ biosensor (Figures 1.1B and D) has been initiated.

In the subsequent sections of this Chapter, an overview is given of the history, development and applications of fluorescent proteins [21].

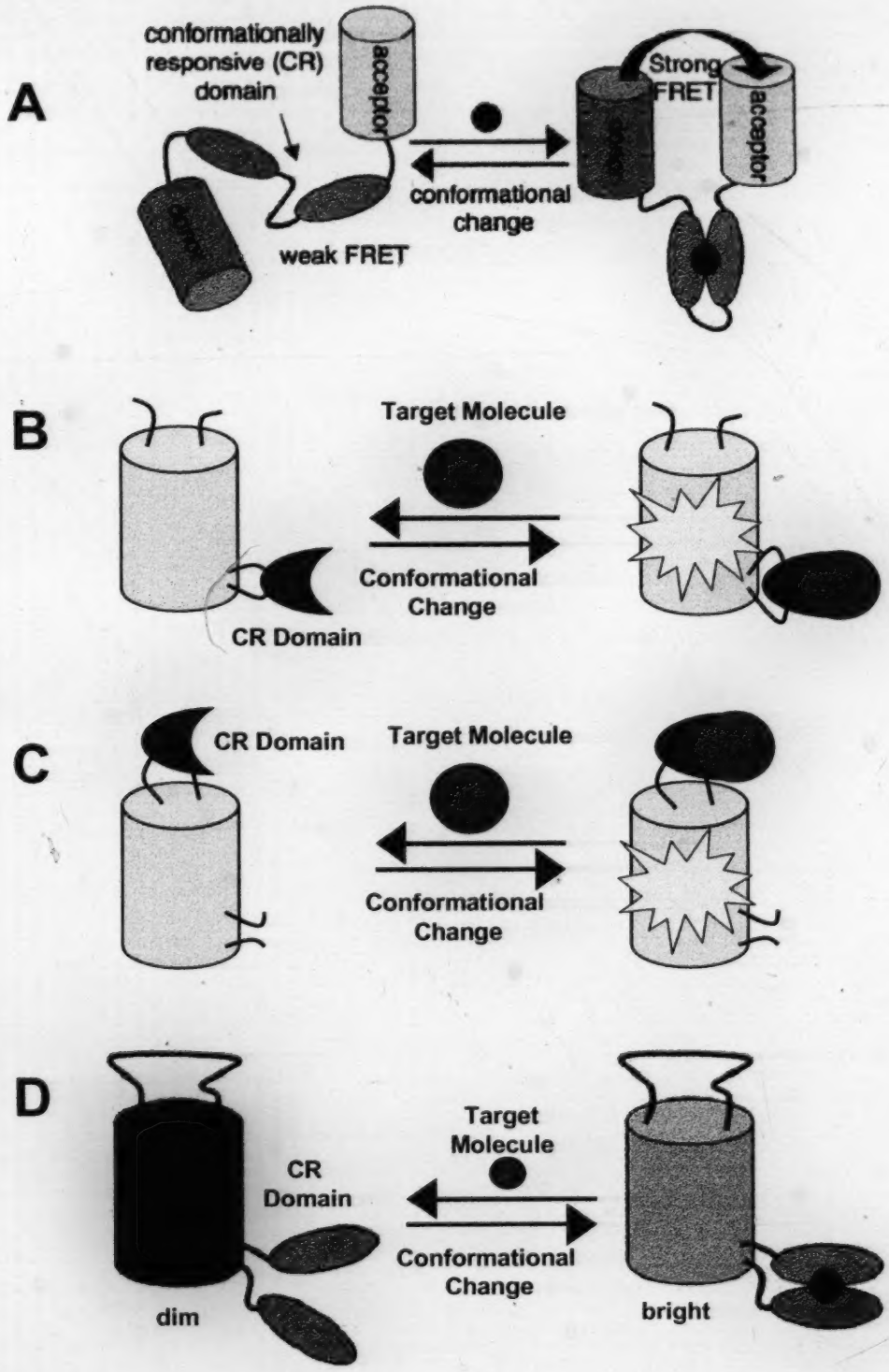

A
conformationally responsive (CR) domain
acceptor
donor
weak FRET
conformational change
Strong FRET
acceptor
donor
B
Target Molecule
Conformational Change
CR Domain
C
CR Domain
Target Molecule
Conformational Change
D
CR Domain
Target Molecule
Conformational Change
dim
bright

**Figure 1.1** FP-based biosensor design. A. FRET-type biosensor. B. Single FP-type biosensor with an inserted conformationally responsive domain within FP. C. Single FP-type biosensor with an inserted conformationally responsive domain within cpFP. D. Single FP-type biosensor with a fused conformationally responsive domain at the new N- and C- termini of cpFP.

## 1.1 *AEQUOREA* GREEN FLUORESCENT PROTEINS (GFP)

### 1.1.1 DISCOVERY OF *AEQUOREA* GFP

Bioluminescence is the emission of visible light by living organisms; an ability that is shared by a wide variety of organisms ranging from bacteria to insects to fish. *Aequorea victoria* jellyfish is one of those organisms capable of emitting light. Its bioluminescence is due to the presence of aequorin (a protein that is sensitive to calcium) and GFP. The protein partners, aequorin and GFP, were discovered and extracted from *Aequorea* jellyfish in 1961 by Shimomura *et al.* [22].

*Aequorea* jellyfish looks like a transparent, hemispherical umbrella, with its mouth at the underside of the body (Figures 1.2A and B). The light-emitting organs, consisting of about 200 tiny granules, are distributed evenly along the edge of the umbrella, making a full circle (Figure 1.2C). A "ring" of bright green light can be seen in the darkness (Figure 1.2D) when a specimen of the jellyfish is placed in dilute potassium chloride (KCl) [23]. The "ring" can be cut off and 20-30 rings can be homogenized to make a turbid liquid, called the "squeezate" [22]. Aequorin and GFP were extracted from the squeezate as described by Shimomura

in 1962: "A protein giving solutions that look slightly greenish in sunlight though only yellowish under tungsten lights, and exhibiting a very bright, greenish fluorescence in the ultraviolet of a Mineralite, has also been isolated from squeezates." [22]. Progress towards the complete characterization of GFP occurred quite slowly: the emission spectrum was published in 1962 [24]; the monomer molecular weight was obtained in 1978 [2[illegible]; the structure of the chromophore was proposed in 1979 [27]; and crystal structure, absorbance spectrum and fluorescence quantum yield were reported in 1994 [25]. Only when GFP was cloned in 1992 [5], did it start to become a popular tool used in cell biology, and initiate a whole new era in biotechnology.

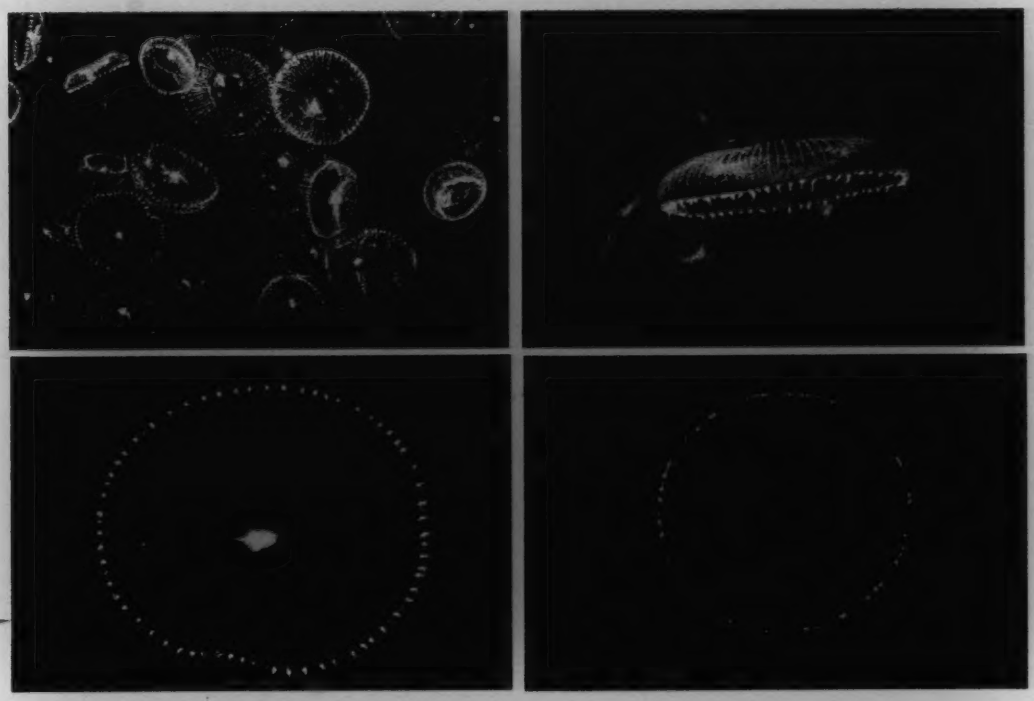


**Figure 1.2** Mid-summer specimens of *Aequorea Jellyfish* photographed in natural environment (A, B & C) and in seawater supplemented with KCl in a darkroom (D) at the University of Washington's Friday Harbor Laboratories [23].

### 1.1.2 STRUCTURE OF AEQUOREA GFP

#### 1.1.2.1 Primary Sequence from Cloning

The sequence of wild-type (wt) *Aequorea* GFP [5] is given in Figure 1.3.

#### 1.1.2.2 The chromophore structure of GFP

The mature chromophore is a p-hydroxybenzylideneimidazolinone [5, 28] formed from residues 65–67, which are Ser-Tyr-Gly in the native protein. Figure 1.4 shows the commonly cited mechanism [4, 29, 30] for chromophore formation. First, GFP folds into a nearly native conformation, and then the imidazolinone is formed by nucleophilic attack of the amide nitrogen of Gly67 on the carbonyl of residue Ser65, followed by dehydration. Finally, molecular oxygen dehydrogenates the α-β bond of residue Tyr66 to put its aromatic group into conjugation with the imidazolinone. Only at this stage does the chromophore acquire visible absorbance and fluorescence.

#### 1.1.2.3 Crystal Structures

GFP was first crystallized in 1974 [25] and diffraction patterns first reported in 1988 [31]. In 1996, the wild type GFP crystal structure was described simultaneously by Yang *et al.* [32] and Ormo *et al.* [33]. GFP resembles a cylinder with a diameter of about 24 Å and a length of 42 Å. Accordingly, it is often referred to as a "β-can" structure. It has a well-packed structure that lacks

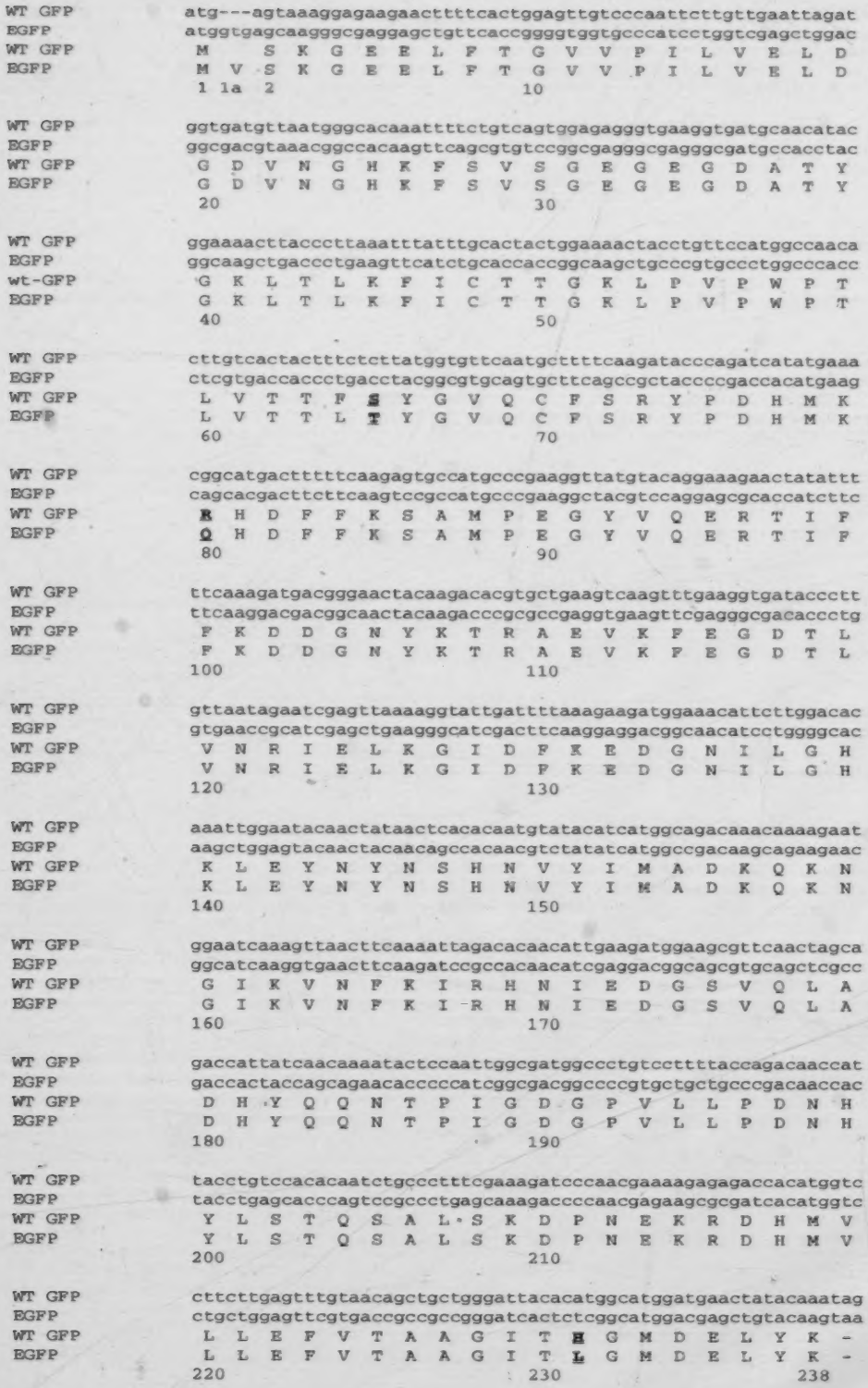

WT GFP atg---agtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagat
EGFP atggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggac
WT GFP M S K G E E L F T G V V P I L V E L D
EGFP M V S K G E E L F T G V V P I L V E L D
1 1a 2 10
WT GFP ggtgatgttaatgggcacaaattttctgtcagtggagagggtgaaggtgatgcaacatac
EGFP ggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctac
WT GFP G D V N G H K F S V S G E G E G D A T Y
EGFP G D V N G H K F S V S G E G E G D A T Y
20 30
WT GFP ggaaaacttacccttaaatttatttgcactactggaaaactacctgttccatggccaaca
EGFP ggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccacc
wt-GFP G K L T L K F I C T T G K L P V P W P T
EGFP G K L T L K F I C T T G K L P V P W P T
40 50
WT GFP cttgtcactactttctcttatggtgttcaatgcttttcaagatacccagatcatatgaaa
EGFP ctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaag
WT GFP L V T T F S Y G V Q C F S R Y P D H M K
EGFP L V T T L T Y G V Q C F S R Y P D H M K
60 70
WT GFP cggcatgactttttcaagagtgccatgcccgaaggttatgtacaggaaagaactatattt
EGFP cagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttc
WT GFP R H D F F K S A M P E G Y V Q E R T I F
EGFP Q H D F F K S A M P E G Y V Q E R T I F
80 90
WT GFP ttcaaagatgacgggaactacaagacacgtgctgaagtcaagtttgaaggtgatacccttt
EGFP ttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctg
WT GFP F K D D G N Y K T R A E V K F E G D T L
EGFP F K D D G N Y K T R A E V K F E G D T L
100 110
WT GFP gttaatagaatcgagttaaaaggtattgattttaaagaagatggaaacattcttggacac
EGFP gtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcac
WT GFP V N R I E L K G I D F K E D G N I L G H
EGFP V N R I E L K G I D F K E D G N I L G H
120 130
WT GFP aaattggaatacaactataactcacacaatgtatacatcatggcagacaaacaaaagaat
EGFP aagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaac
WT GFP K L E Y N Y N S H N V Y I M A D K Q K N
EGFP K L E Y N Y N S H N V Y I M A D K Q K N
140 150
WT GFP ggaatcaaagttaacttcaaaattagacacaacattgaagatggaagcgttcaactagca
EGFP ggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgcc
WT GFP G I K V N F K I R H N I E D G S V Q L A
EGFP G I K V N F K I R H N I E D G S V Q L A
160 170
WT GFP gaccattatcaacaaaatactccaattggcgatggccctgtccttttaccagacaaccat
EGFP gaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccac
WT GFP D H Y Q Q N T P I G D G P V L L P D N H
EGFP D H Y Q Q N T P I G D G P V L L P D N H
180 190
WT GFP tacctgtccacacaatctgccctttcgaaagatcccaacgaaaagagagaccacatggtc
EGFP tacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtc
WT GFP Y L S T Q S A L S K D P N E K R D H M V
EGFP Y L S T Q S A L S K D P N E K R D H M V
200 210
WT GFP cttcttgagtttgtaacagctgctgggattacacatggcatggatgaactatacaaatag
EGFP ctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagtaa
WT GFP L L E F V T A A G I T H G M D E L Y K -
EGFP L L E F V T A A G I T L G M D E L Y K -
220 230 238

**Figure 1.3** GFP sequences. (Lines 1 and 2) WT GFP and EGFP gene, respectively, Clontech Laboratories. (Line 3 and 4) WT GFP and EGFP amino acid sequence, respectively. (Lines 5) Numbering of amino acids of both EGFP and WT GFP. The inserted Val in EGFP is numbered 1a to maintain correspondence with the WT GFP numbering.

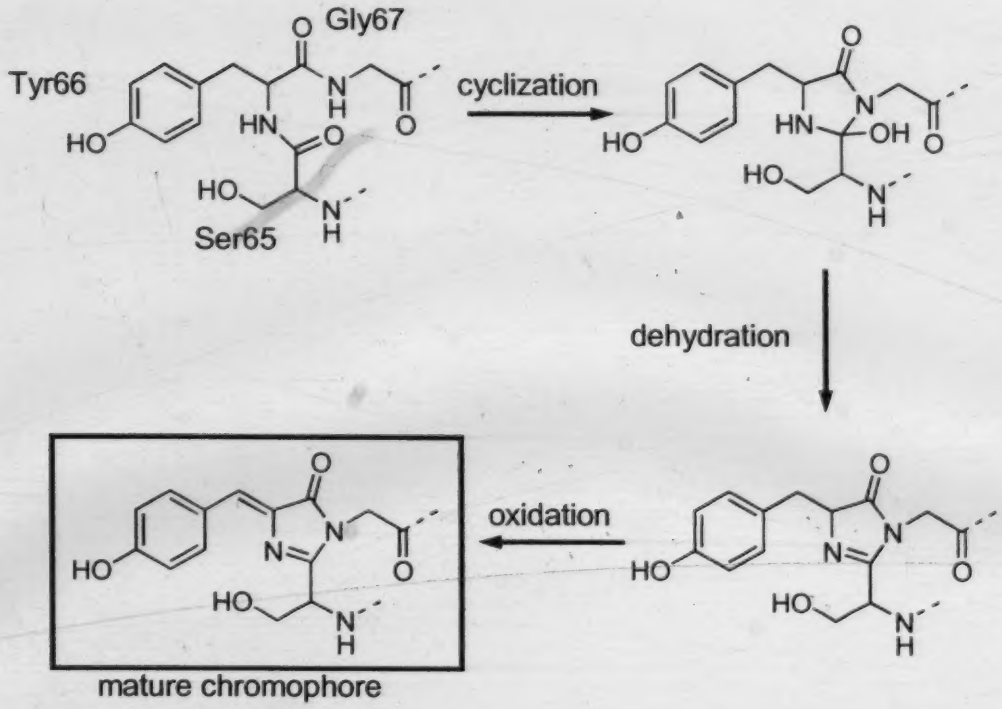



**Figure 1.4** Mechanism proposed by Cubitt et al. [29] for the intramolecular biosynthesis of the GFP chromophore.

any large extended loop domains [32]. The cylindrical fold of the protein is made up of an 11-stranded β-sheet. The continuous β-sheet wraps around an α-helix running up the axis of the cylinder (Figure 1.5). The chromophore is formed from three residues of the central α-helix and is approximately positioned at the geometric center of the β-can [32, 34]. Small sections of α-helix also form caps

on the top and bottom of the β-can. The cavity that contains the chromophore is lined by a number of charged residues and also contains four water molecules that are important in establishing a hydrogen-bonding network around the chromophore (Figure 1.6). This network might be critical in establishing the spectroscopic and photochemical properties of GFP. Particularly important are Gln69, Arg96, His148, Thr203, Ser205, and Glu222. There is a slight separation between the strands of the β-can along the face of the β-can to which the chomophore faces (comprising residues 145-150 and 164-169) [35]. Residues 1 and 230–238 were too disordered to be resolved; these regions correspond closely to the maximal amino- and carboxyl-terminal deletions that still permit fluorescence to develop [36].

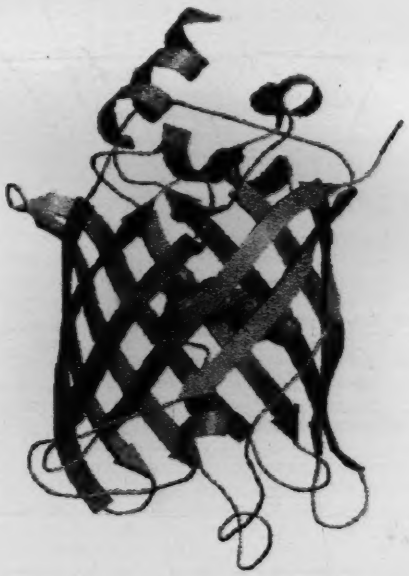


**Figure 1.5** A ribbon representation of the three-dimensional structure of GFP[33], showing 11 β-strands forming a hollow cylinder through which is threaded a helix bearing the chromophore, shown in Van der Waals Radii representation.

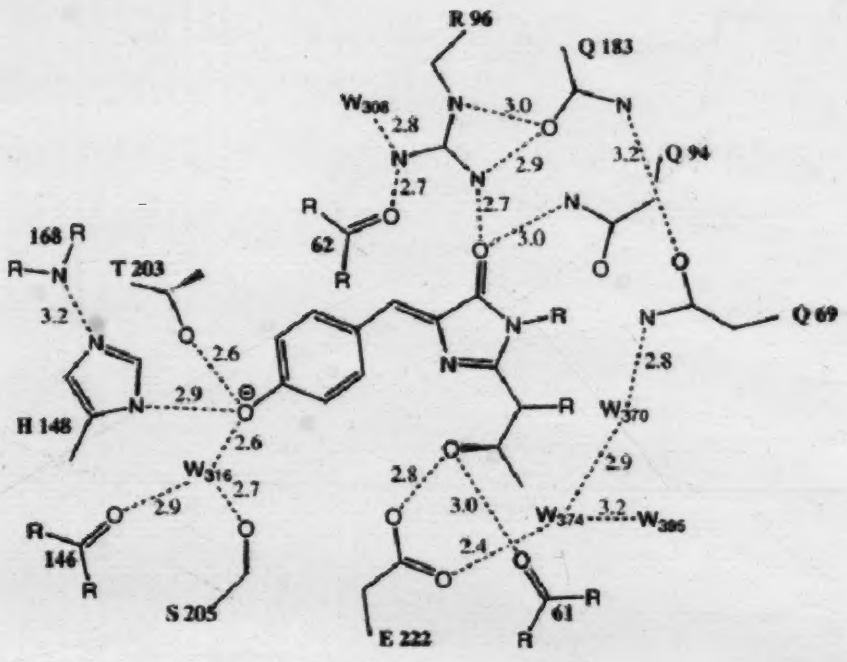



**Figure 1.6** Amino acid side chains, main chain carbonyls and amides, and solvent waters in the immediate vicinity of the chromophore of S65T GFP [33]. Side chains are labeled with the one-letter code for the amino acid and the residue number. Main chain groups are labeled with the residue number. Water oxygens are denoted by W and the corresponding residue number in the PDB file. Hydrogen bonds are shown as dotted lines labeled with the distance between the heteroatoms in angstroms. From reference [8].

## 1.2 ANTHOZOAN FLUORESCENT PROTEINS

### 1.2.1 DISCOVERY OF FLUORESCENT PROTEINS IN CORALS

Since the introduction of *Aequorea victoria* green fluorescent protein (GFP) into cell biology and biotechnology practices [8], looking for other FPs with emission colors other than green has been an active area of research. The

availability of FPs with different colors would permit multicolor tracking of multiple fusion proteins simultaneously or create pairs of donors and acceptors for fluorescence resonance energy transfer (FRET). Although protein engineering efforts successfully produced GFP variants with blue, cyan, and yellowish-green fluorescence, no *Aequorea* GFP variants with true yellow, orange or red fluorescence have ever been engineered [33].

The search for novel FPs was continued in other bioluminescence systems. Soon, GFP homologues were found in several bioluminescent organisms, including *Renilla reniformis* [37], *Hydrozoa* and *Anthozoa* [38]. A breakthrough in the quest for novel colors of FPs was the cloning by Matz *et al.* [12] of six *Anthozoan* fluorescent proteins from reef corals. This work demonstrated that FPs can be present in nonbioluminenscent organisms and their fluorescent properties can demonstrate a dramatic red shift. All the six new coral proteins shared the same β-can fold first observed in GFP. One of the new FPs, zFP538 emitted at yellow (538 nm) wavelength and one, drFP583 from *Discosoma sp.,* emitted at red (583 nm) wavelength. The red-shifted fluorescent proteins allowed tracking two markers simultaneously because they offered the convenience of minimizing background auto-fluorescence. Soon other GFP-like proteins were found in other corals [12, 39, 40], which might contain an almost limitless inventory of fluorescent proteins because of the abundance of coral fluorescence and the impressive variety of colors [41].

### 1.2.2 PROPERTIES OF DSRED

Because of the special properties of GFP, it has rapidly become a standard tool for investigating a variety of cellular activities, and has served as a model system for understanding spectral tuning in chromophoric proteins. The GFP-like proteins from reef corals *Anthozoa* display some new properties of the fluorescent protein family, most notably a wide variety of spectral hues. However, there were also drawbacks including oligomerization to form tetramers and extremely slow maturation. In the following sections, the properties of drFP583 [12] from *Discosoma sp.* "red" (commercially termed DsRed) will be introduced as the representative of the *Anthozoan* fluorescent proteins, because mCherry used in this thesis work was derived from it.

#### 1.2.2.1 Spectroscopic properties

Once fully maturated, the fluorescence emission spectrum of DsRed [12] features a very broad peak at 583 nm (spectral width about 50–60 nm), whereas the excitation spectrum has a major peak at 558 nm and a minor peak around 500 nm. The emission spectrum of maturated DsRed fluorescent protein still has a green region, termed the green state, which has proven problematic for multiple labeling experiments with other green fluorescent proteins because of the spectral overlap.

#### 1.2.2.2 Oligomerizaion

DsRed (Figure 1.9), like other coral GFP-like proteins [42-47], was found to be an obligate tetramer even at nanomolar concentration [48-53]. Each DsRed monomer contacts the two adjacent protein molecules by two chemically distinct interfaces: one hydrophobic and the other hydrophilic. The hydrophobic interface includes a central cluster of closely packed hydrophobic residues surrounded by polar side chains. The hydrophilic interface contains many salt bridges and hydrogen bonds between polar residues and buried water molecules and also includes an unusual "clasp" formed by several C-terminal residues of each monomer.

The oligomeric structure makes DsRed have greater stability than monomeric FPs at elevated temperatures and under other physical factors *in vitro* [54, 55], and a longer intracellular lifespan *in vivo* [12, 54]. The distinct second interface maximizes the photostability and thermotolerance of DsRed [56]. The tetrameric DsRed is four- to five-fold more resistant to photobleaching than *A. victoria* GFP and tenfold more resistant than its monomeric mutant mRFP1 [10].

Despite these advantages, oligomerization of DsRed is a big problem for its use as a fluorescent reporter. DsRed is an obligate tetramer and can form large protein aggregates *in vivo* and *in vitro* [44, 57, 58]. The aggregation may impede targeting to cell compartments and interfere with investigation of protein interactions. In contrast to *Aequorea* GFP, which has been successfully used to tag hundreds of proteins, DsRed conjugates have proven much less successful and are often toxic.

#### 1.2.2.3 Slow maturation

The last property and also a serious problem of DsRed is slow and incomplete maturation. The half-life of DsRed is more than 24 hours at room temperature and it keeps residual green fluorescence [48, 59].

### 1.2.3 STRUCTURE OF DSRED

#### 1.2.3.1 Primary Sequence from Cloning

The amino acid sequence of DsRed identity with *Aequorea victoria* GFP is very low (~25%) [60]. However, the overall β-can fold structure of GFP [32, 33] is highly conserved in DsRed (Figure 1.9) [50, 51, 61]. Figure 1.7 shows the sequences of WT DsRed and mCherry derived from WT DsRed (1.4.1.2).

#### 1.2.3.2 Chromophore Structures in *Anthozoa* GFP Homologs

The chromophore of DsRed is formed by a cyclization reaction from amino acids Gln66, Tyr67 and Gly68 (Figure 1.8). The first steps of chromophore formation likely involve nucleophilic attack of the Tyr67 amide nitrogen on the Gln66 backbone carbonyl followed by elimination of a water molecule. These first two steps (cyclization and dehydration steps) generate the imidazolone ring. Oxidation step 1 generates a green fluorescing structure by addition of $O_2$ + $H^+$ and elimination of $H_2O$. Oxidation step 2 involving the amide nitrogen of Phe65 extends the conjugation of the intermediate green chromophore to produce the red fluorescing chromophore. The αC of Gln66 has a planar $sp^2$ conformation

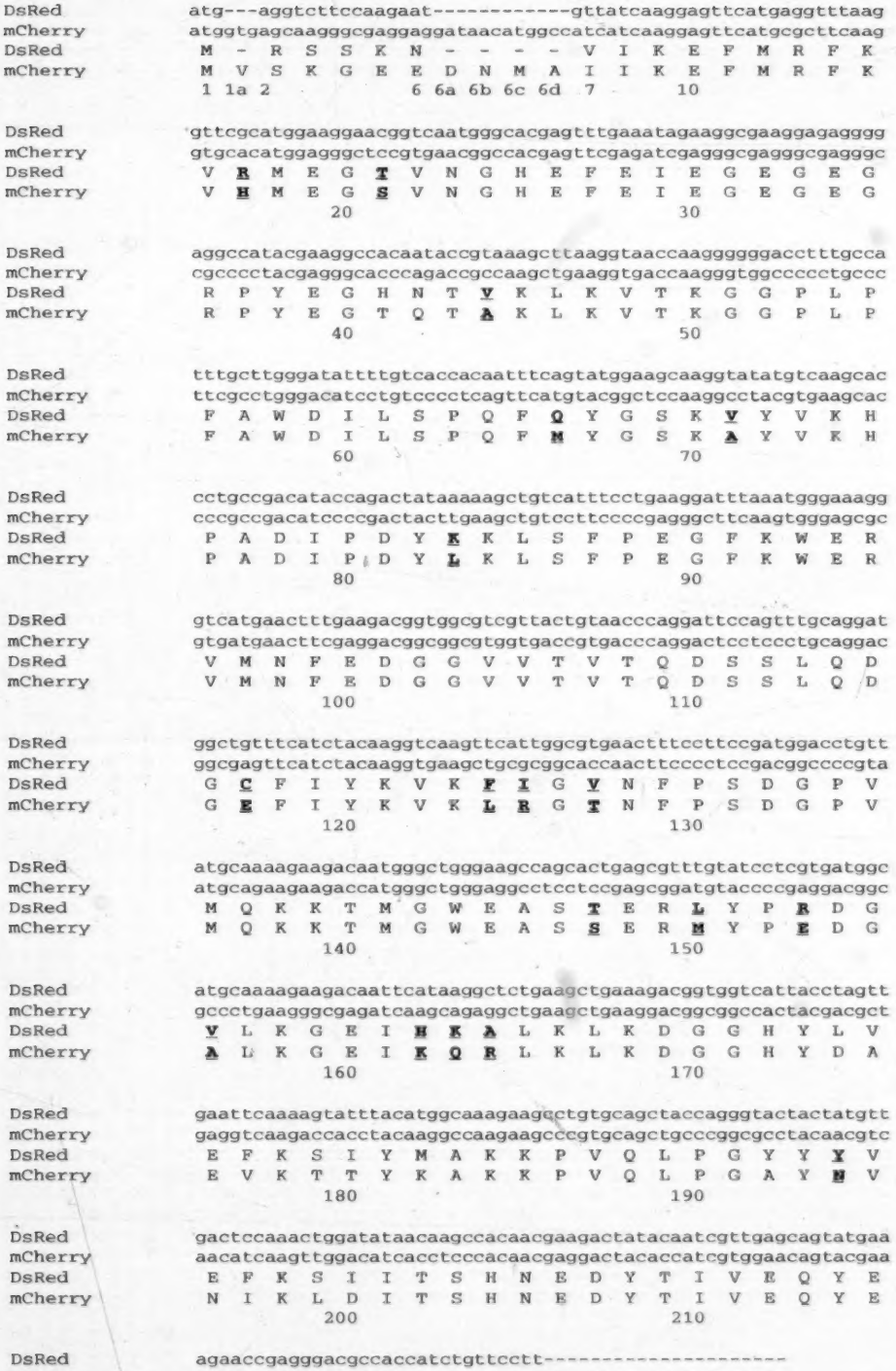

DsRed atg---aggtcttccaagaat----------gttatcaaggagttcatgaggtttaag
mCherry atggtgagcaagggcgaggaggataacatggccatcatcaaggagttcatgcgcttcaag
DsRed M - R S S K N - - - - V I K E F M R F K
mCherry M V S K G E E D N M A I I K E F M R F K
1 1a 2 6 6a 6b 6c 6d 7 10
DsRed gttcgcatggaaggaacggtcaatgggcacgagtttgaaatagaaggcgaaggagagggg
mCherry gtgcacatggagggctccgtgaacggccacgagttcgagatcgagggcgagggcgagggc
DsRed V R M E G T V N G H E F E I E G E G E G
mCherry V H M E G S V N G H E F E I E G E G E G
20 30
DsRed aggccatacgaaggccacaataccgtaaagcttaaggtaaccaagggggacctttgcca
mCherry cgccctacgagggcacccagaccgccaagctgaaggtgaccaagggtggccccctgccc
DsRed R P Y E G H N T V K L K V T K G G P L P
mCherry R P Y E G T Q T A K L K V T K G G P L P
40 50
DsRed tttgcttgggatattttgtcaccacaatttcagtatggaagcaaggtatatgtcaagcac
mCherry ttcgcctgggacatcctgtcccctcagttcatgtacggctccaaggcctacgtgaagcac
DsRed F A W D I L S P Q F Q Y G S K V Y V K H
mCherry F A W D I L S P Q F M Y G S K A Y V K H
60 70
DsRed cctgccgacataccagactataaaaagctgtcatttcctgaaggatttaaatgggaaagg
mCherry cccgccgacatccccgactacttgaagctgtccttccccgagggcttcaagtgggagcgc
DsRed P A D I P D Y K K L S F P E G F K W E R
mCherry P A D I P D Y L K L S F P E G F K W E R
80 90
DsRed gtcatgaactttgaagacggtggcgtcgttactgtaacccaggattccagtttgcaggat
mCherry gtgatgaacttcgaggacggcggcgtggtgaccgtgacccaggactcctccctgcaggac
DsRed V M N F E D G G V V T V T Q D S S L Q D
mCherry V M N F E D G G V V T V T Q D S S L Q D
100 110
DsRed ggctgtttcatctacaaggtcaagttcattggcgtgaactttccttccgatggacctgtt
mCherry ggcgagttcatctacaaggtgaagctgcgcggcaccaacttcccctccgacggccccgta
DsRed G C F I Y K V K F I G V N F P S D G P V
mCherry G E F I Y K V K L R G T N F P S D G P V
120 130
DsRed atgcaaaagaagacaatgggctgggaagccagcactgagcgtttgtatcctcgtgatggc
mCherry atgcagaagaagaccatgggctgggaggcctcctccgagcggatgtaccccgaggacggc
DsRed M Q K K T M G W E A S T E R L Y P R D G
mCherry M Q K K T M G W E A S S E R M Y P E D G
140 150
DsRed atgcaaaagaagacaattcataaggctctgaagctgaaagacggtggtcattacctagtt
mCherry gccctgaagggcgagatcaagcagaggctgaagctgaaggacggcggccactacgacgct
DsRed V L K G E I H K A L K L K D G G H Y L V
mCherry A L K G E I K Q R L K L K D G G H Y D A
160 170
DsRed gaattcaaaagtatttacatggcaaagaagcctgtgcagctaccagggtactactatgtt
mCherry gaggtcaagaccacctacaaggccaagaagcccgtgcagctgcccggcgcctacaacgtc
DsRed E F K S I Y M A K K P V Q L P G Y Y Y V
mCherry E V K T T Y K A K K P V Q L P G A Y N V
180 190
DsRed gactccaaactggatataacaagccacaacgaagactatacaatcgttgagcagtatgaa
mCherry aacatcaagttggacatcacctcccacaacgaggactacaccatcgtggaacagtacgaa
DsRed E F K S I I T S H N E D Y T I V E Q Y E
mCherry N I K L D I T S H N E D Y T I V E Q Y E
200 210
DsRed agaaccgagggacgccaccatctgttcctt----------
mCherry cgcgccgagggccgccactccaccggcggcatggacgagctgtacaagtaa
DsRed E F K S I H H L F L
mCherry R A E G R H S T G G M D E L Y K -
220 230

**Figure 1.7** Sequences. (Lines 1 and 2) WT DsRed and mCherry gene, respectively, Clontech Laboratories. (Line 3 and 4) WT DsRed and mCherry amino acid sequence, respectively. (Lines 5) Numbering of amino acids of both WT DsRed and mCherry.

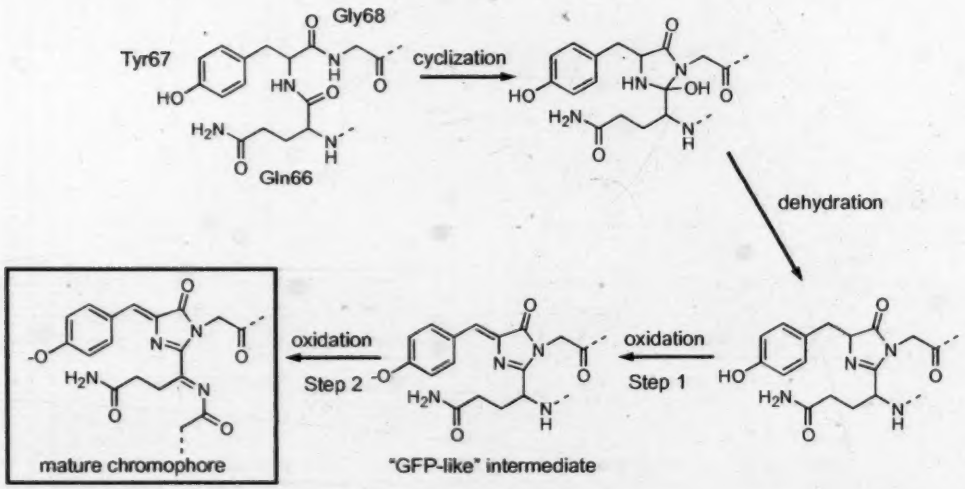



**Figure 1.8** Scheme for the formation of mature red chromophore of DsRed by conversion from the green intermediate by oxidation. Proposed by Gross et al. [62].

extending the π-electron system through the additional imine moiety. The adjoining carbonyl of Phe65 is $sp^2$ hybridized, too, but its plane is perpendicular to the chromophore and thus may not fully contribute to a larger resonance system [50, 51, 62]. In the mature tetrameric protein only half of the chromophores are oxidized in the second step, so really there is a mixture of green and red fluorescent proteins.

#### 1.2.3.3 The crystal structure of *Anthozoa* proteins

As mentioned above, DsRed is tetrameric [50, 51] (Figure 1.9) and each of the monomers are identical. Like GFP, each monomer of DsRed is an 11-stranded β-can with a central α-helix; the DsRed chromophore rests in the middle of an α-helix in the center of the β-can; and the top and bottom of each β-can is sealed off from bulk solvent by protein atoms that protect the chromophore. The majority of the backbone structural variation between DsRed and GFP occurs in loop regions that form the ends of the β-can structure. The distances between the indicated chromophore pairs of A–B and C–D are 22 Å and the distances of A–C and B–D are 38 Å, and 43 Å between the two diagonally related chromophores. The angular orientations of these potential FRET pairs in DsRed are 21°, 47°, and 41°, respectively [50].

## 1.3 APPLICATION IN FLUORESCENCE MICROSCOPY

Fluorescence illumination and observation is the most rapidly expanding microscopy technique employed today [63], both in the medical and biological sciences, because it allows researchers to image live cells with high spatial and temporal resolution using relatively inexpensive and generally accessible equipment. The molecular prerequisite for undertaking fluorescence imaging of a specific protein in a live cell is that the protein of interest be conjugated to a fluorescent probe. By necessity, synthetic organic dyes were the only practical fluorescent probes for imaging of biological macromolecule up until the 1990s.

During the 1990s, fluorescent proteins rapidly replaced synthetic dyes for live cell imaging applications. The reason for the extreme popularity of FPs is that these proteins provide genetic encoding of strong visible fluorescence of a wide range of colors [64] that span almost the entire visible light spectrum.

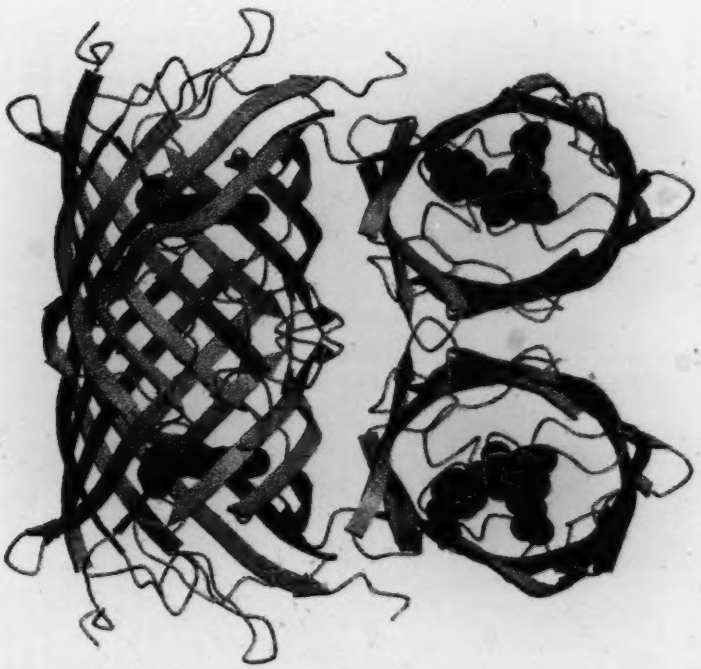


**Figure 1.9** Stereoview of the three-dimensional structure of DsRed showing 11 β-strands forming a hollow cylinder through which is threaded a helix bearing the chromophore; shown in Van der Waals Radii representation.

### 1.3.1 INTRODUCTION TO FLUORESCENCE MICROSCOPY

#### 1.3.1.1 Fluorescence

The phenomenon of fluorescence was first observed by British scientist Sir George G. Stokes in the middle of the nineteenth century. Fluorescence is the ability of some atoms and molecules to absorb light at a particular wavelength and subsequently emit photons with longer wavelength (known as Stokes Shift [66]) after a short-lived interval, termed the fluorescence lifetime. The process of phosphorescence occurs in a manner similar to fluorescence, but it has a much longer excited state lifetime [65]. Fluorescence is the result of a three-stage process that occurs in certain molecules called fluorophores or fluorescent dyes, as illustrated by the Jablonski diagram (Figure 1.10) [65].

When illuminated with ultraviolet or visible light, common fluorophores are usually excited to higher vibrational levels of the first (S(1)) or second (S(2)) singlet excited energy state from the ground energy state (S0), which is known as the absorption transition. Two examples are presented in Figure 1.10: one is from S(0) = 0 to S(2) = 3, the other is from S(0) = 1 to S(1) = 5. Immediately following absorption of a photon, several processes will occur, but the most likely will be internal conversion or vibrational relaxation in which electrons are relaxed to the lowest vibrational energy level of the first excited state (S(1) = 0). Generally, an excited molecule exists in the lowest excited singlet state (S(1) = 0) for periods on the order of nanoseconds before finally relaxing to the ground state. Fluorescence is a process in which the relaxation from this long-lived state is accompanied by emission of a photon. Most fluorophores can repeat the excitation and emission cycle many hundreds to thousands of times in response to incident illumination

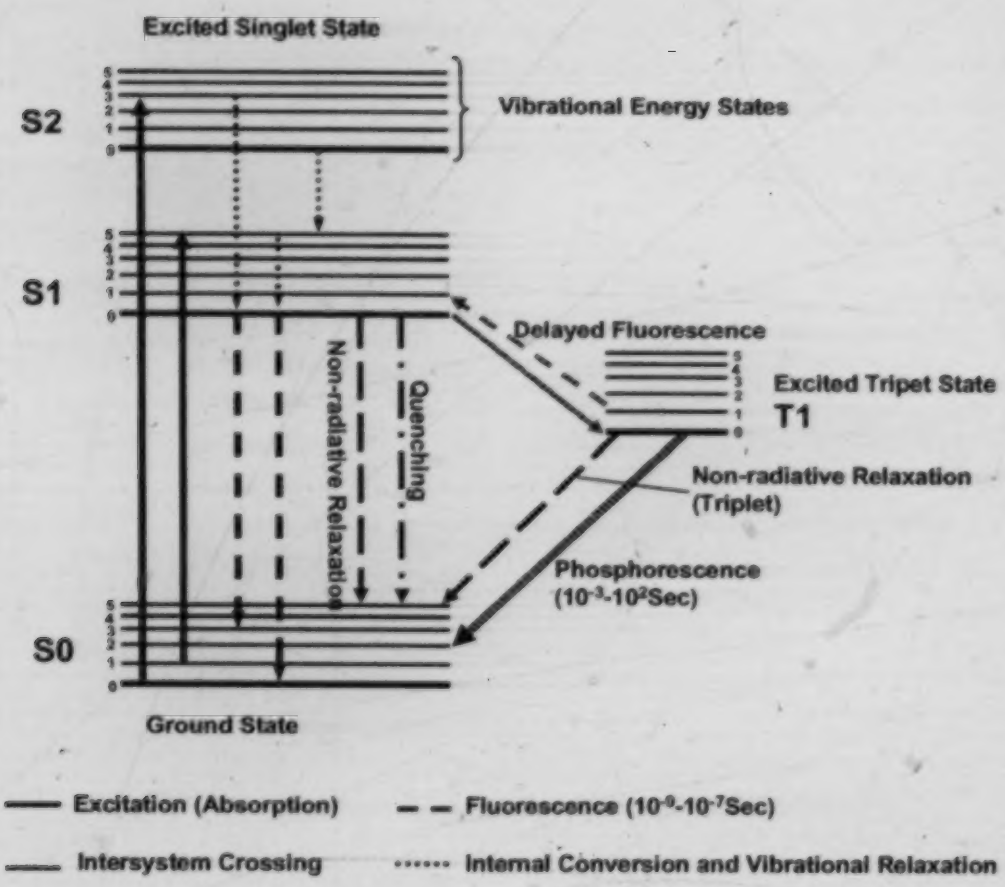



**Figure 1.10** Jablonski diagram.

before the highly reactive excited state molecule is photobleached, which results in the destruction of fluorescence. Several other relaxation pathways also compete with the fluorescence emission process: 1) the excited state energy can be dissipated non-radiatively as heat; 2) the excited fluorophore can collide with another molecule to transfer energy in a second type of non-radiative process called quenching; and 3) intersystem crossing to the lowest excited triplet state can occur, which rarely occurs. However, if it occurs, it ultimately results either in

emission of a photon through phosphorescence or a transition back to the excited singlet state that yields delayed fluorescence.

### 1.3.1.2 Fluorescence Microscopy

Fluorescence microscopy is a rapidly expanding and invaluable tool of investigation. The earliest fluorescence microscope developed was a type of transmitted light microscope (Figure 1.11 A) in which excitation light passed through a filter onto the specimen and fluorescence emission with a significant amount of the excitation illumination were gathered by the objective after passage through a second filter. The images obtained with the transmitted light microscope suffered from poor resolution and very low brightness level because the intensity of the excitation light was usually several orders of magnitude greater than the fluorescence emission.

In the late 1920s, transmitted light fluorescence microscopes were replaced by epi-illumination fluorescence microscopes (Figure 1.11B). The big improvement of the epi-illumination fluorescence microscopy is achieved by dichroic mirrors as beam-splitters used to separate the emission light and excitation light in order to obtain either an image of the emission without excessive background illumination, or a measurement of the fluorescence emission without background "noise". Therefore, epi-illumination fluorescence microscopes produce images with a significantly greater level of brightness than

transmitted light microscopes. Nowadays, nearly all fluorescence microscopes that use lamp light sources use epi-illumination.

In epi-illumination fluorescence microscopy, the UV and visible light is selected through an excitation filter and then focused by a condenser lens system. Then the excitation light passes through the dichroic mirror and then illuminates the fluorescent species. The sample can either be fluorescing in its natural form (termed primary or autofluorescence) or treated with fluorescing chemicals capable of fluorescing (known as secondary fluorescence). By means of the dichroic mirror the emission light and the excitation light are separated and the emission light goes through the emission filter. The emission filter is required to select specifically the emission wavelength of the light emitted from the sample and to remove traces of excitation light. The filtered emission light is focused by a condenser lens system before it goes to a detector.

### 1.3.2 FLUORESCENT PROTEINS VS. BIOLUMINESCENT PROTEINS VS. SMALL MOLECULE FLUOROPHORES FOR CALCIUM SENSING IN LIVE CELLS.

The imaging mode of fluorescence microscopy relies heavily on fluorescence which is provided by fluorescent fluorophores (fluorescent probes). Many small molecule fluorophores (dyes) have been constructed by synthesis of aromatic organic molecules designed to bind with a biological macromolecule (such as a protein or nucleic acid), or to localize within a specific structural region (such as the cytoskeleton or mitochondria), or to monitor dynamic processes. Fluorescent

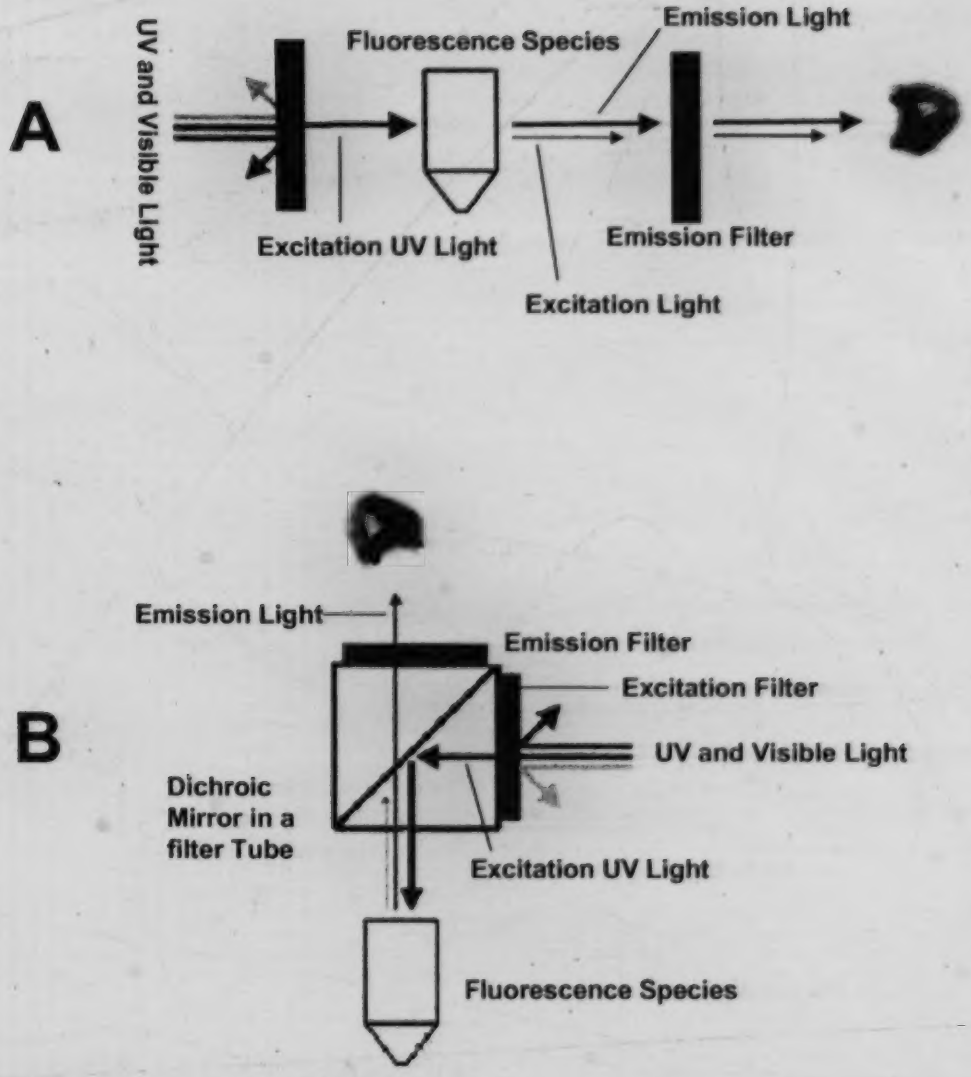



**Figure 1.11** Diagrams of transmitted light microscopy (A) and epi-illumination fluorescence microscopy (B).

proteins have had a very important role as fluorescent probes used in fluorescence microscopy studies since their cloning [5].

In fluorescence microscopy studies, cytosolic and organellar free $Ca^{2+}$ concentration are among the most important and dynamic intracellular signals. For example, $Ca^{2+}$ signals can regulate the transmit signals within cells. $Ca^{2+}$ regulates specific cellular processes, such as muscle contraction. Calmodulin, normally present in the cytoplasm, binds calcium ions and changes conformation. This change in conformation allows calmodulin to bind many different target proteins, one of which is Myosin Light Chain (MLCK). The interaction of CaM and MLCK allows smooth muscle contraction. However, the $Ca^{2+}$ signals are not able to be detected visually. So, $Ca^{2+}$ signals are most often measured using $Ca^{2+}$ sensitive fluorescent dyes, such as Fura-2 (Figure 1.12 A) [67] or Fura Red (Figure 1.12 B) [68], or the bioluminescent protein aequorin [69-71], or fluorescent proteins.

Fura-2 [67] has excitation at 300-350 nm and green emission at around 510 nm. Upon binding $Ca^{2+}$, fura-2 exhibits an absorption shift that can be observed by scanning the excitation spectrum. The spectrum of the Fura Red indicator [68], a kind of fura-2 analog, is more red shifted with visible-wavelength excitation at 450–500 nm and a very long-wavelength emission maximum at ~660 nm. This minimizes interference from autofluorescence in tissues and biological fluids [72]. Upon binding $Ca^{2+}$, the fluorescence of the Fura Red [68] indicator decreases.

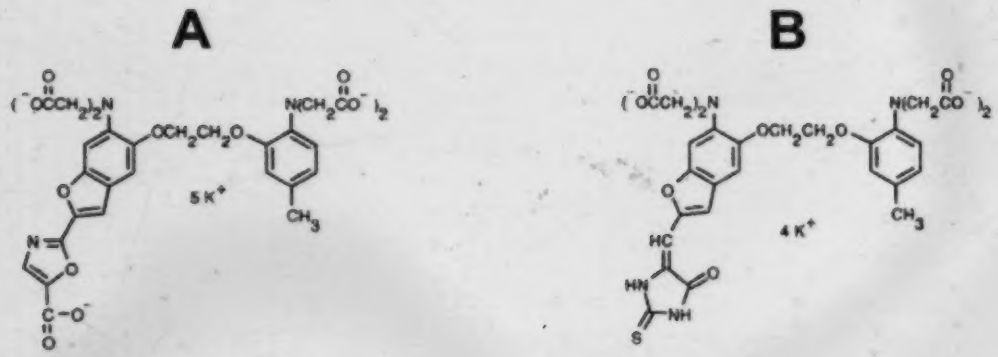



**Figure 1.12** The structure of fura-2, pentapotassium salt (A) and Fura Red, tetrapotassium salt (B).

The aequorin complex [69-71] comprises a 22,000 Da apoaequorin protein, molecular oxygen and the luminophore, coelenterazine. When three $Ca^{2+}$ ions bind to the complex, coelenterazine is oxidized to coelenteramide, with a concomitant release of carbon dioxide and blue light.

Many kinds of indicator, such as Cameleons, Camgaroos and Pericams, based on fluorescent proteins [21] have been developed to monitor $Ca^{2+}$ concentrations. This will be described in section 1.6.

Each probe has its own advantages and disadvantages listed in Table 1.1. The small molecule fluorophores are easy to load by microinjection in isolated cells but hard or impossible to load in thicker tissues. They also make it easy to image single cells, but they are difficult to target precisely to specific subcellular locations. By contrast, aequorin is easy to target, but very difficult to use to image single cells which require specialized photon-counting systems [70], because the

photon intensity is extremely low and its luminescence produces less than 1 photon/molecule. Moreover, aequoein requires the incorporation of a cofactor, coelenterazine, and gene transfer.

Comparing them, the fluorescent protein indicators have very good targetability, which can make it easy to target specific intracellular sites (Figure 1.13 B). They also make it easy to image single cells because they are bright. Moreover, unlike the small molecule fluorophores and aequorin, the fluorescent proteins can be introduced into cells by DNA transfection of the protein chimeras (Figure 1.13 A) without using any invasive methods like protein microinjection, but gene transfer is required. In addition, the luminescence of aequorin can be detectable many hours to days after loading and the luminescence of fluorescent proteins can be detectable even longer. However, the small fluorophores gradually leak out of cells.

In addition, fluorescent protein indicators have many advantages over others. They can be observed at low light intensities for long time periods with minimal photobleaching. Fluorescent proteins from jellyfish or corals exhibit a variety of absorption and emission characteristics across the entire visible spectral region, which make it possible to develop probe combinations for simultaneous observation of two or more distinct fluorescent proteins in a single organism. Furthermore, the $Ca^{2+}$ indicators can detect $Ca^{2+}$ in a wide concentration range by mutation of the CaM moiety [15, 73]. Also the cost of synthesizing the small molecule fluorophores is higher than the fluorescent protein indicators because the

latter just need DNA transfection once the indicator is made. Of course, the fluorescent protein indicators still have some other disadvantages. For example, EYFP is sensitive to changes in pH, which interferes with calcium sensors.

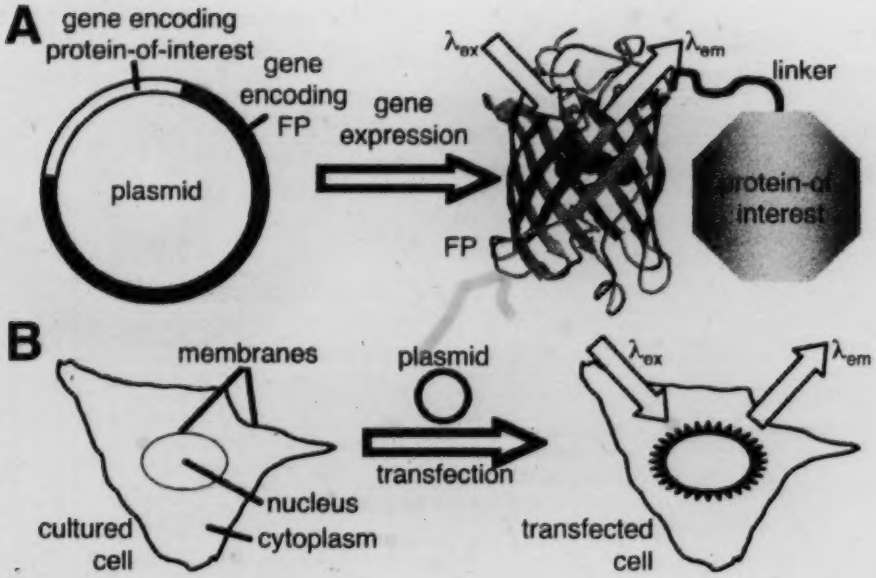



**Figure 1.13** A. Genetic construction of a chimera. B. Illustrative example in which the fusion protein targets to the nuclear membrane.

**Table 1.1** Comparison of the characteristics of the fluorescent probes: small molecules, aequoin and fluorescent proteins. The more +, the more advantages.

| Characteristics | Small molecules | aequorin | FPs |
|---|---|---|---|
| Load | + | ++ | +++ |
| Targeting | + | ++ | +++ |
| Single-cell imaging | +++ | + | +++ |
| Long time measurement | + | ++ | +++ |
| Ease of making sensors | + | + | ++ |
| Ease of use | ++ | + | + |
| Brightness | +++ | + | ++ |
| Photostability | + | + | ++ |
| Selection of colors | +++ | + | +++ |
| Cost | ++ | + | +++ |
| Dynamic range | +++ | ++ | +++ |
| pH | ++ | ++ | + |
| Cofactor | +++ | + | +++ |

## 1.4 ENGINEERING OF FLUORESCENT PROTEINS

### 1.4.1 ENGINEERING NEW COLORS

In the past several years, FPs have been further expanded through the engineering of blue, cyan, green, yellow, orange, red, and far-red variants, spanning almost the entire visible light spectrum [74], making multicolor imaging of fusion proteins and FRET measurements in living cells possible [8-12]. The extensive mutagenesis of *Aequorea* GFP provides fluorescent proteins ranging in color from blue to yellow [8], and the mutagenesis from *Anthozoa* corals provides fluorescent proteins with longer wavelength [56].

### 1.4.1.1 Aequorea GFP Variants

The *Aequorea* GFP variants were classified by chromophore type into seven classes by Tsien in 1998 [8]: class 1, wild-type mixture of neutral phenol and anionic phenolate; class 2, phenolate anion (EGFP); class 3, neutral phenol; class 4, phenolate anion with stacked π-electron system (YFP); class 5, indole in chromophore (CFP); class 6, imidazole in Chromophore (BFP) ; and class 7, phenyl in chromophore. Each class has distinct excitation and emission spectra. Classes 1–4 are derived from polypeptides with Tyr at position 66, while classes 5–7 result from Trp, His, and Phe at position 66.

**Green Fluorescent Proteins (GFPs)**

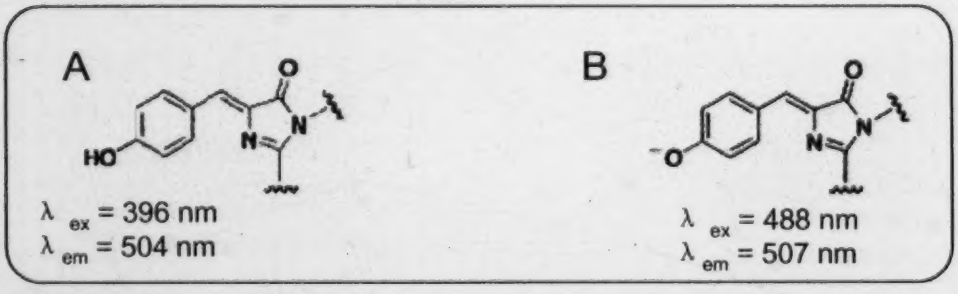



**Figure 1.14** The chromophore of WT *Aequorea* GFP (A) and EGFP (B). Abbreviations: $\lambda_{ex}$, excitation wavelength in nm and $\lambda_{em}$ emission wavelength, which will be used in Figures 1.15 to 1.19.

The wild type *Aequorea* GFP has bimodal absorption at 395 and 475 nm [22]. Enhanced green fluorescent protein (EGFP) [75][76, 77], made by the point mutation S65T, has a well defined absorption profile with a single peak at 484 nm. In addition to EGFP, there are other mutations to make other green-emitting

variants (in the range of approximately 500 to 525 nm): Emerald variant [77] with the best photostability, brightness and folding ability; Sapphire [78] with a large Stokes shift of over 100 nm; and T-Sapphire [78] with better folding ability than Sapphire. The mechanism of WT GFP chromophore formation is shown in Figure 1.4 and the chromophore and the spectral properties of WT GFP and EGFP are shown in Figure 1.14.

**Blue Fluorescent Proteins (BFPs)**

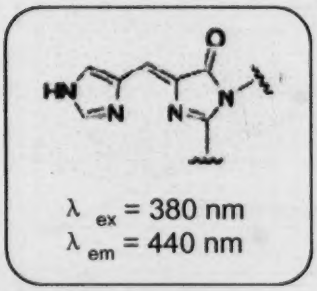



**Figure 1.15** The chromophore of the original BFP.

Blue fluorescent proteins (BFPs), with emission wavelength maxima ranging from 440 to 470 nm, were first obtained by site-directed mutagenesis efforts to obtain a Y66H mutant (Figure1.15). The original BFP exhibits a broad absorption band in the UV centered close to 380 nm, and has only about 15 to 20 pecentt of the parent GFP brightness due to a low quantum yield. Several additional mutations (i.e. Y64L, Y66H and Y145F) led to enhanced BFP (EBFP) versions that are still only 25 percent as bright as the enhanced green variants and display limited photostability compared to many other fluorescent proteins [76, 77].

**Cyan Fluorescent Proteins (CFPs)**

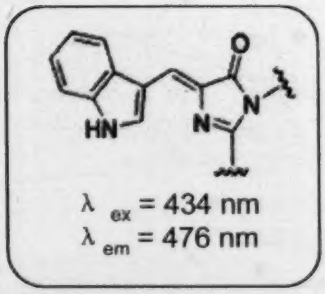



**Figure 1.16** The chromophore of CFP or Cerulean.

Cyan fluorescent protein (CFP; emission wavelength maxima ranging from approximately 470 to 500 nm) was discovered by introducing the Y66W mutation into GFP [4]. This protein displays a very broad fluorescence emission spectral profile. subsequent maturation resulted in an enhanced version (ECFP, F64L and S65T) with greater brightness and photostability [77], which is also a good FRET pair of YFP. The Cerulean variant [79] was rationally engineered by site-directed mutagenesis of ECFP to yield a higher extinction coefficient and improved quantum yield. Recent developments have resulted in CyPet [80]( Cyan fluorescent Protein for energy transfer) which was developed to enhance the cyan and yellow pairing for FRET.

**Yellow Fluorescent Proteins (YFPs)**

Yellow fluorescent proteins (YFPs, ranging in emission wavelength maxima from approximately 525 to 555 nm), the furthest red-shifted of all the *Aequorea* mutants, resulted from rational introduction of a π-stacked aromatic side chain (T203Y) next to the chromophore [77, 81]. The mutation at position 203T to

aromatic residues (His, Trp, Phe and Tyr (the mutants called 10C) – in order of least to greatest red-shifting potential) makes the excitation and emission spectra 20 nm red-shifted [33] [77]. Further mutations at residues Q69K [77] and Q69M (Citrine) [82] cause an additional 1 to 2 nm red shift. Another novel point mutation (F46L) from circularly permuted GFP (cpGFP) produced the brighter variant, Venus [83, 84]. Further CyPet, developed YPet (YFP for energy transfer) [80], the brightest YFP, which acted as a good FRET pair of CFP and also demonstrates very good photostability.

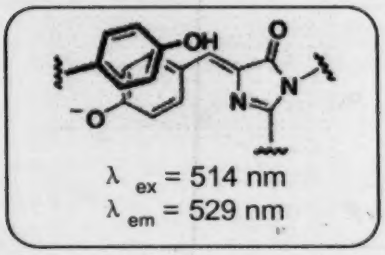



**Figure 1.17** The chromophore of YFP.

### 1.4.1.2 *Anthozoan* Proteins

The *Anthozoan* proteins [12] provides a rich new source of GFP-like fluorescent proteins [56]: GFPs, YFPs, and RFPs and non-fluorescent chromoproteins (CPs), from orange to blue. The far red-shifted fluorescent proteins that were discovered, such as the most well-studied red protein - DsRed [48], are important because they make multicolor labeling more feasible and also provide better imaging reagents for expression in heterologous systems. The longer the light wavelength, the higher the sensitivity and the greater the efficiency of fluorescence detection in whole body imaging.

**DsRed**

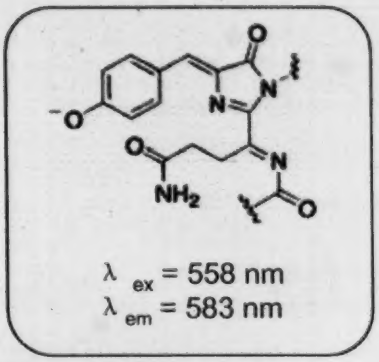



**Figure 1.18** The chromophore of DsRed.

As mentioned previously, the tetrameric DsRed has many FP undesirable problems: slow and incomplete maturation, aggregation and oligomerization, that have been already described and greatly limit its applications. In 2002, a dimeric DsRed [10] was created by mutation (I126R) and additional mutagenesis resultinge in a bright dimeric DsRed. More importantly, a true monomeric protein, named mRFP1, was genetated by multiple steps of random and site-directed mutagenesis (in total 33 substitutions compared to DsRed).

**mRFP1 and its Variants – The "Fruit" Series of Fluorescent Proteins**

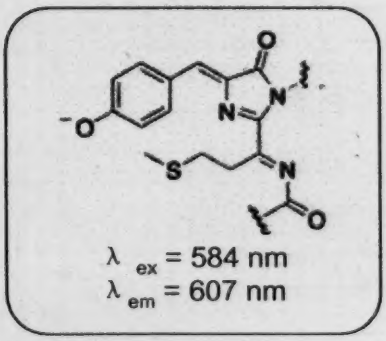



**Figure 1.19** The chromophore of mRFP1.

mRFP1 possesses several improved characteristics relative to DsRed: more red shift, absence of residual green fluorescence, and very fast maturation [10]. However, mRFP1 had reduced fluorescence emission compared to DsRed and photobleached very quickly, rendering it less than ideal. Through

extensive mutagenesis, a series of fluorescent proteins have been made and named in honor of common fruits that bear similar colors. Roger Y. Tsien [85] said: "One of the reasons we name the proteins after fruits is to remind people that there is no 'best fruit' in the grocery store." The fluorescent features of these "fruits" are: increased extinction coefficients, quantum yields, and photostability and extension of the color palette [13, 14, 86]. Typically, through directed mutagenesis targeting of the Q66 and Y67 chromophore residues [14], fluorescent proteins with emission maxima ranging from 560 to 610 nm have been made: mHoneydew ($\lambda_{em}$ = 553nm), mBanana ($\lambda_{em}$ = 562 nm), mOrange ($\lambda_{em}$ = 581nm), tdTomato (a tandem dimer, $\lambda_{em}$ = 581 nm ), mTangerine ($\lambda_{em}$ = 585 nm), mStrawberry ($\lambda_{em}$ = 596 nm), and mCherry ($\lambda_{em}$ = 610 nm). Iterative somatic hypermutation (SHM) has yielded two fluorescent proteins mRaspberry ($\lambda_{em}$ = 625 nm) and mPlum ($\lambda_{em}$ = 649 nm) [87]. Of these proteins, mCherry has less brightness values (approximately 50 percent of EGFP), but is far more photostable and is the best probe choice for long-term imaging experiments. Further, mCherry2, a folding-optimized version of mCherry was discovered through more mutagenesis (Nathan Shaner, personal communication). These fruit fluorescent proteins should be useful in combination with other fluorescent proteins for multicolor imaging experiments and as a biosensor FRET partner with green and yellow proteins.

## 1.5 ENGINEERING "STRUCTURAL" VARIANTS

### 1.5.1 FUSIONS

FPs have had an important role in visualizing intracellular organelles or other subcellular structures in living cells because colorful FPs can make proteins or organelles dynamic in a real time. The most successful application of FPs is fusing FPs to host proteins without changing their normal functions and making them fluorescent so that their localization and fate can be monitored (Figure 1.13). For example, the fluorophore-conjugated protein, GFP fused with the microtubule binding region of human ensconsin, was expressed in living tissue to monitor cytoskeletal dynamics [88].

### 1.5.2 INSERTIONS

The perfect barrel-like structures of FPs give us an image of a protein that appears to be rather monolithic, seemingly without much tolerance for gross rearrangements and insertions. However, a six-residue peptide (FKTRHN) had been inserted into CFP at position 145 without destroying its fluorescence [17]. This suggested that the insertion of an entire protein with a conformationally responsive domain within FPs could be used to develop a new generation of molecular biosensors, called Camgaroo (Figure 1.1B.). Soon, many breakpoints were found in FPs [19, 78, 90-97], where peptides or an entire protein with a conformationally responsive domain could be inserted without quenching the fluorescence of FPs.

### 1.5.3 CIRCULAR PERMUTATIONS

Circular permutation is an alternative method of protein engineering that entails the manipulation of protein sequences to a much greater extent [98].

Conceptually, a circularly permuted variant is a protein in which the original N- and C- termini are joined by a flexible linker, and new N- and C- termini are introduced somewhere else within the protein. In general, two requirements must be fulfilled for a successful circular permutation experiment: the termini of the corresponding protein must be in close proximity, and the introduction of the new termini may not affect folding of the protein to its functional tertiary structure [98]. In a thought experiment (Figure 1.20), the concept can be considered as follows: a wild-type protein in which the termini are in close proximity is circularized by linking the N- and C- terminal ends directly, or via a short linker peptide. The circular protein is subsequently cleaved at another position in the sequence, generating new termini that are again in close proximity.

Since the first synthesis of a circularly permuted protein in bovine pancreatic trypsin inhibitor (BPTI) [99], circular permutation studies have extended to many proteins with the N- and C-termini in close proximity, including the FPs. In the three-dimensional crystal structure of FPs, which consists of an 11-stranded β-barrel, the p-hydroxybenzylideneimidazolidone chromophore is located within a helical segment in the center of the barrel, and is entirely shielded from the solvent [64]. At first glance, the well-conserved β-barrel structures of FPs might seem to indicate that rearrangements, insertions, or deletions within the main polypeptide backbone would prevent fluorescence. However, the successful insertion of a six-residue peptide within GFP [17] gave researchers the idea that circular permutations of FPs (cpFPs) might be possible. The breakpoint within

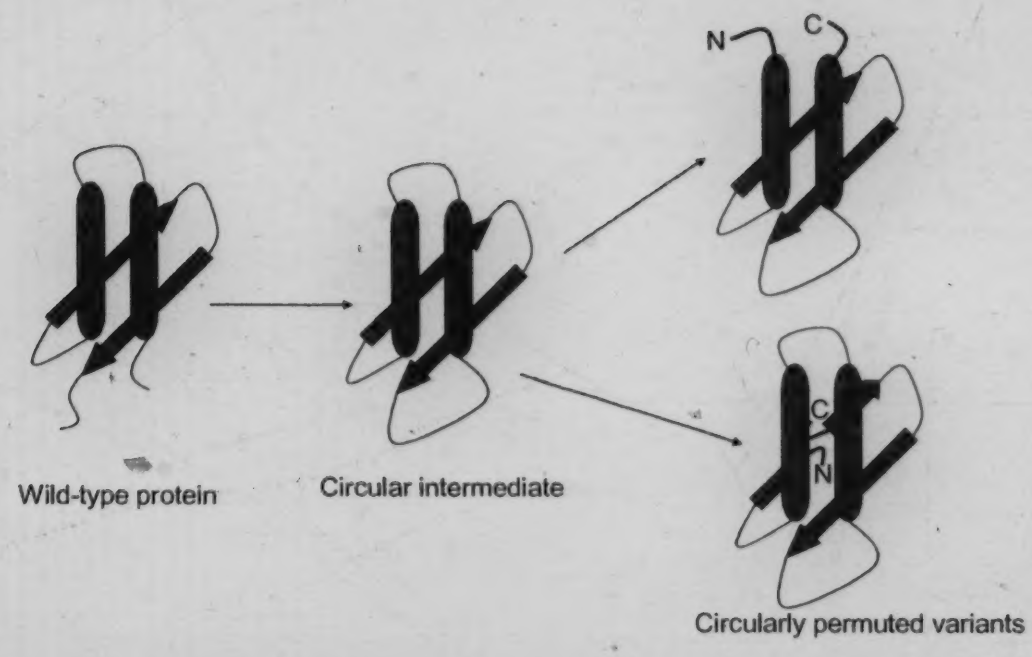



**Figure 1.20** The principle of circular permutation of a polypeptide chain. The N- and C- termini of the wild-type protein are connected, generating a circular protein intermediate. The protein backbone is then cleaved at a different position, yielding a circularly permuted variant of the protein with novel N- and C- termini.

GFP was cleaved and generated new C- and N- termini, and the original termini were circularized by linking the N- and C- terminal ends directly, or by a short linker peptide (Figure 1.21).

Green [17, 78, 90], enhanced green [92] and yellow fluorescent [94, 96, 97] proteins have been genetically engineered to create fluorescent circular permutations of the original sequences (abbreviated cpGFP, cpEGFP and cpYFP). The properties of the circular permutation variants are somewhat altered with

respect to the standard fluorescent proteins and these new designs offer a unique opportunity to generate an original class of localization probes and physiological indicators[9]. Such permutation increases the flexibility and optical responsiveness to stresses that are applied to the new termini [9].

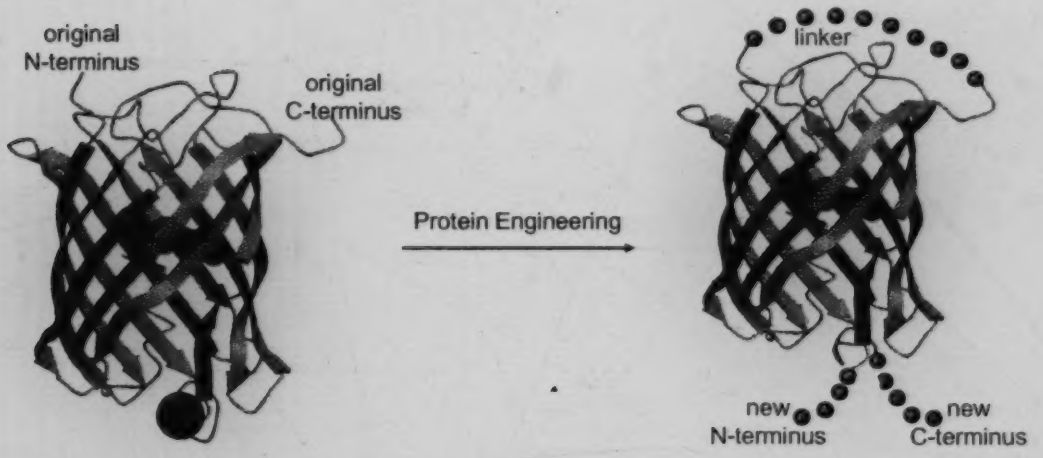



**Figure 1.21** The general idea of engineering circularly permuted mCherry created from insertion variants.

## 1.6 ENGINEERING SENSORS

The rigid shell in FPs surrounding the chromophore not only protects it from photobleaching but also hinders environmental sensitivity. However, by mutagenesis, FPs still can be sensitive to their environments to make biosensors.

### 1.6.1 FRET SENSORS

The most general way to make biochemically sensitive FPs is to exploit fluorescence resonance energy transfer (FRET) between FPs with different colors

[8]. FRET is a quantum mechanical phenomenon and is involves a distance dependent dipole-dipole interaction between two fluorophores. FRET takes place when two fluorophores are in molecular proximity (<100 Å apart) and one fluorophore (the donor) transfers the excitation energy directly to a second fluorophore (the acceptor). This process happens without the emission of a photon that would normally occur in the absence of the acceptor and is at the expense of the emission from the donor (Figure 1.22).

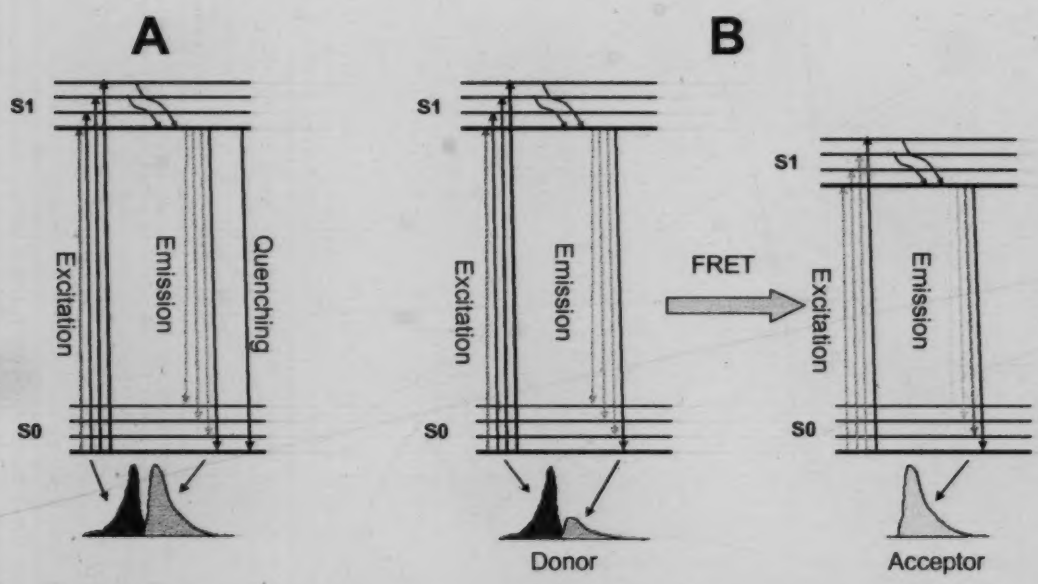



**Figure 1.22** Jablonski diagram to show the principle of (A) fluorescence (B) FRET.

Because FRET depends on the distance between these two fluorophores, any biochemical signal that changes the distance between the fluorophores or

relative orientation of their transition dipoles will modulate the efficiency of FRET. Based on FRET of FPs, many sensors [15, 100-102] have been made. Among these sensors are the $Ca^{2+}$ sensors known as Cameleons [15, 100] (Figure 1.1A). Cameleons consist of a CFP and YFP that are linked by calmodulin (CaM), a protein that changes its shape in the presence of calcium, and its binding peptide, M13, from skeletal muscle myosin light chain kinase. The binding of $Ca^{2+}$ to the CaM causes it to associate with the M13 peptide. This leads to a conformational change in the molecule, bringing the two FPs into closer molecular proximity, which in turn increases the efficiency of FRET between the two FPs [15]. This design has been used to construct generic kinase activity biosensors that can be used to address questions such as: '*When a cell is stimulated with an agonist, where and when in the cell does the activity of kinase X exhibit the greatest increase?*' The identity of kinase X is arbitrary and can be changed to suit the specific biological question by a simple genetic modification of the biosensor [9, 15, 16].

### 1.6.2 SINGLE FP SENSORS

Single FP-type biosensors consist of an FP with a conformationally responsive domain (such as calmodulin) [17-20] (Figures 1.1B and C). The single-FP-type biosensor offers many advantages over FRET-type biosensors (e.g. half the size and typically much greater fluorescent response), but is intrinsically less modular and less generalizable. A dramatic example of the use of this type of

reporter was the 2-photon imaging of odor perception in *Drosophila* (fruit fly) brain [103].

### 1.6.2.1 Camgaroo indicators

The first Camgaroo type of indicator (Figure 1.1B) was a BLIP (β-lactamase-inhibitory protein) sensor [18], in which β-lactamase (a protein that binds to BLIP) was genetically inserted into GFP between the 172Q and 173D sites. Later, the first $Ca^{2+}$ Camgaroo sensor [19] was made by inserting calmodulin into YFP between the 145N and 146I sites, and has proven useful in measuring calcium inside mitochondria and mushroom bodies in the brain of Drosophila [104]. In this Camgaroo, fluorescence intensity increases sevenfold on binding of $Ca^{2+}$. Soon, Camgaroo $Ca^{2+}$ sensors were made in EGFP[20] and EYFP [19]. Similarly, $Zn^{2+}$ sensors were made by inserting the zinc finger motif from zif268 into YFP 145 [17], and BFP [105]. Hydrogen peroxide sensor [96] was constructed by inserting $H_2O_2$-sensitive protein OxyR into YFP. A Camgaroo would be ideal if robust, selective sensors of such simplicity could be generated for cellular analytes [8].

### 1.6.2.2 Pericams

Circularly permuted FP (cpFP) was inserted between calmodulin (CaM) and M13 to yield $Ca^{2+}$ sensors that have been termed Pericams (Figure 1.1C). The inclusion of M13 increases the apparent $Ca^{2+}$ affinity of the CaM by allowing the

formation of ternary complexes, so that these molecules are more sensitive than Camgaroos to small elevations in physiological levels of $Ca^{2+}$. Some Pericam variants shift their excitation wavelengths on binding of $Ca^{2+}$, as opposed to just increasing fluorescence, which thereby enables ratiometric observation [19, 93, 106].

### 1.6.3 TRANSLOCATION SENSORS

Translocation sensors are a type of FP-based sensor in which the fluorescent intensity or color of the FP does not change. Rather, a biochemical event is detected by a change in the localization of FP-protein fusion protein in a live cell. A typical sort of localization change might involve moving from the cytoplasm to the plasma membrane. To make translocation-type sensors, the FPs are fused to minimal protein domains that interact specifically with small molecule messengers to monitor the cellular localization and transient production of the messengers. The pleckstrin homology domain of the AKT protein kinase (or protein kinase B), tagged with the green fluorescent protein (PHAKT-GFP), was expressed in neutrophils to monitor polarization of chemoattractant receptor [89].

## 1.7 OBJECTIVES OF THE THESIS

The objective of this work was to make the mCherry $Ca^{2+}$ sensor of either the Camgaroo-type or Pericam-type. By random insertion, a five-residue peptide was inserted into mCherry at position 183 without destroying its fluorescence. Based on this insertion, the circular permutation of mCherry was carried out and it

remained fluorescent. This work is a crucial first step towards the creation of single-FP red fluorescent $Ca^{2+}$ sensors.

## 1.8 LITERATURE CITED


(1) Fox, D. L., and Pantin, C. F. A. *Biol. Rev.* **1944**, *19*, 121-133.

(2) Cheesman, D. F., Lee, W. L., and Zagalsky P. F. *Biol. Rev.* **1967**, *42*, 132-160.

(3) Zimmer, M. *Glowing genes: a revolution in biotechnology*; Prometheus Books: Amherst, N.Y., 2005.

(4) Heim, R.; Prasher, D. C.; Tsien, R. Y. *Proc. Natl. Acad.Sci. U. S. A.* **1994**, *91*, 12501-12504.

(5) Prasher, D. C.; Eckenrode, V. K.; Ward, W. W.; Prendergast, F. G.; Cormier, M. J. *Gene* **1992**, *111*, 229-233.

(6) Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C. *Science* **1994**, *263*, 802-805.

(7) Inouye, S.; Tsuji, F. I. *FEBS Lett.* **1994**, *341*, 277-280.

(8) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509-544.

(9) Zhang, J.; Campbell, R. E.; Ting, A. Y.; Tsien, R. Y. *Nat. Rev. Mol. Cell. Biol.* **2002**, *3*, 906-918.

(10) Campbell, R. E.; Tour, O.; Palmer, A. E.; Steinbach, P. A.; Baird, G. S.; Zacharias, D. A.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 7877-7882.

(11) Matz, M. V.; Lukyanov, K. A.; Lukyanov, S. A. *Bioessays* **2002**, *24*, 953-959.

(12) Matz, M. V.; Fradkov, A. F.; Labas, Y. A.; Savitsky, A. P.; Zaraisky, A. G.; Markelov, M. L.; Lukyanov, S. A. *Nat. Biotechnol.* **1999**, *17*, 969-973.

(13) Wang, L.; Jackson, W. C.; Steinbach, P. A.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16745-16749.

(14) Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N.; Palmer, A. E.; Tsien, R. Y. *Nat. Biotechnol.* **2004**, *22*, 1567-1572.

(15) Miyawaki, A.; Llopis, J.; Heim, R.; McCaffery, J. M.; Adams, J. A.; Ikura, M.; Tsien, R. Y. *Nature* **1997**, *388*, 882-887.

(16) Ting, A. Y.; Kain, K. H.; Klemke, R. L.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 15003-15008.

(17) Baird, G. S.; Zacharias, D. A.; Tsien, R. Y. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11241-11246.

(18) Doi, N.; Yanagawa, H. *FEBS Lett.* **1999**, *453*, 305-307.

(19) Nagai, T.; Sawano, A.; Park, E. S.; Miyawaki, A. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 3197-3202.

(20) Nakai, J.; Ohkura, M.; Imoto, K. *Nat. Biotechnol.* **2001**, *19*, 137-141.

(21) Zacharias, D. A.; Tsien, R. Y. *Methods Biochem. Anal.* **2006**, *47*, 83-120.

(22) Shimomura, O.; Johnson, F. H.; Saiga, Y. *J. Cell Comp. Physiol.* **1962**, *59*, 223-239.

(23) Shimomura, O. *Methods Biochem Anal.* **2006**, *47*, 1-13.

(24) Johnson, F. H.; Shimomura, O.; Saiga, Y. *J. Cell Comp. Physiol.* **1962**, *59*, 265-272.

(25) Morise, H.; Shimomura, O.; Johnson, F. H.; Winant, J. *Biochemistry* **1974**, *13*, 2656-2662.

(26) Prendergast, F. G.; Mann, K. G. *Biochemistry* **1978**, *17*, 3448-3453.

(27) Shimomura, O. *FEBS Lett.* **1979**, *104*, 220-222.

(28) Cody, C. W.; Prasher, D. C.; Westler, W. M.; Prendergast, F. G.; Ward, W. W. *Biochemistry* **1993**, *32*, 1212-1218.

(29) Cubitt, A. B.; Heim, R.; Adams, S. R.; Boyd, A. E.; Gross, L. A.; Tsien, R. Y. *Trends Biochem. Sci.* **1995**, *20*, 448-455.

(30) Reid, B. G.; Flynn, G. C. *Biochemistry* **1997**, *36*, 6786-6791.

(31) Perozzo, M. A.; Ward, K. B.; Thompson, R. B.; Ward, W. W. *J. Biol. Chem.* **1988**, *263*, 7713-7716.

(32) Yang, F.; Moss, L. G.; Phillips, G. N., Jr. *Nat. Biotechnol.* **1996**, *14*, 1246-1251.

(33) Ormo, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J. *Science* **1996**, *273*, 1392-1395.

(34) Phillips, G. N., Jr. *Curr. Opin. Struct. Biol.* **1997**, *7*, 821-827.

(35) Sullivan, K. F.; Kay, S. A.; American Society for Cell Biology. *Green fluorescent proteins*; Academic Press: San Diego, Calif.; London, 1999.

(36) Dopf, J.; Horiagon, T. M. *Gene* **1996**, *173*, 39-44.

(37) Ward, W. W.; Cormier, M. J. *J. Biol. Chem.* **1979**, *254*, 781-788.

(38) Chalfie, M. *Photochem Photobiol* **1995**, *62*, 651-656.

(39) Wiedenmann, J.; Elke, C.; Spindler, K. D.; Funke, W. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 14091-14096.

(40) Dove, S. G. H.-G., O.; Ranganathan, S. *Coral Reefs* **2001**, *19*, 197-204.

(41) Mazel, C. H. *Mar. Ecol. Prog. Ser.* **1995**, *120*, 185-191.

(42) Verkhusha, V.; Akovbian, N.; Efremenko, E.; Varfolomeyev, S.; Vrzheshch, P. *Biochemistry-Moscow* **2001**, *66*, 1342-1351.

(43) Gurskaya, N. G.; Fradkov, A. F.; Terskikh, A.; Matz, M. V.; Labas, Y. A.; Martynov, V. I.; Yanushevich, Y. G.; Lukyanov, K. A.; Lukyanov, S. A. *FEBS Lett.* **2001**, *507*, 16-20.

(44) Yanushevich, Y. G.; Staroverov, D. B.; Savitsky, A. P.; Fradkov, A. F.; Gurskaya, N. G.; Bulina, M. E.; Lukyanov, K. A.; Lukyanov, S. A. *FEBS Lett.* **2002**, *511*, 11-14.

(45) Petersen, J.; Wilmann, P. G.; Beddoe, T.; Oakley, A. J.; Devenish, R. J.; Prescott, M.; Rossjohn, J. *J. Biol. Chem.* **2003**, *278*, 44626-44631.

(46) Prescott, M.; Ling, M.; Beddoe, T.; Oakley, A. J.; Dove, S.; Hoegh-Guldberg, O.; Devenish, R. J.; Rossjohn, J. *Structure* **2003**, *11*, 275-284.

(47) Quillin, M. L.; Anstrom, D. M.; Shu, X.; O'Leary, S.; Kallio, K.; Chudakov, D. M.; Remington, S. J. *Biochemistry* **2005**, *44*, 5774-5787.

(48) Baird, G. S.; Zacharias, D. A.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 11984-11989.

(49) Heikal, A. A.; Hess, S. T.; Baird, G. S.; Tsien, R. Y.; Webb, W. W. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 11996-12001.

(50) Wall, M. A.; Socolich, M.; Ranganathan, R. *Nat. Struct. Biol.* **2000**, *7*, 1133-1138.

(51) Yarbrough, D.; Wachter, R.; Kallio, K.; Matz, M.; Remington, S. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 462-467.

(52) Cotlet, M.; Hofkens, J.; Habuchi, S.; Dirix, G.; Van Guyse, M.; Michiels, J.; Vanderleyden, J.; De Schryver, F. C. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 14398-14403.

(53) Garcia-Parajo, M. F.; Koopman, M.; van Dijk, E. M.; Subramaniam, V.; van Hulst, N. F. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 14392-14397.

(54) Verkhusha, V. V.; Kuznetsova, I. M.; Stepanenko, O. V.; Zaraisky, A. G.; Shavlovsky, M. M.; Turoverov, K. K.; Uversky, V. N. *Biochemistry* **2003**, *42*, 7879-7884.

(55) Verkhusha, V. V.; Pozhitkov, A. E.; Smirnov, S. A.; Borst, J. W.; van Hoek, A.; Klyachko, N. L.; Levashov, A. V.; Visser, A. J. *Biochim. Biophys. Acta.* **2003**, *1622*, 192-195.

(56) Verkhusha, V. V.; Lukyanov, K. A. *Nat. Biotechnol.* **2004**, *22*, 289-296.

(57) Mizuno, H.; Sawano, A.; Eli, P.; Hama, H.; Miyawaki, A. *Biochemistry* **2001**, *40*, 2502-2510.

(58) Jakobs, S.; Subramaniam, V.; Schonle, A.; Jovin, T. M.; Hell, S. W. *FEBS Lett.* **2000**, *479*, 131-135.

(59) Wiehler, J.; von Hummel, J.; Steipe, B. *FEBS Lett.* **2001**, *487*, 384-389.

(60) Lukyanov, K. A.; Chudakov, D. M.; Fradkov, A. F.; Labas, Y. A.; Matz, M. V.; Lukyanov, S. *Methods Biochem. Anal.* **2006**, *47*, 121-138.

(61) Tubbs, J. L.; Tainer, J. A.; Getzoff, E. D. *Biochemistry* **2005**, *44*, 9833-9840.

(62) Gross, L. A.; Baird, G. S.; Hoffman, R. C.; Baldridge, K. K.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 11990-11995.

(63) Axelrod, D. *Fluororescence Microscopy*.

(64) Tsien, R. Y.; Ernst, L.; Waggoner, A. *Fluorophores for Confocal Microscopy: Photophysics and Photochemistry*, 3 ed.; Springer Business Media, LLC.: New York, 2006.

(65) Herman, B., http://microscope.fsu.edu/primer/techniques/fluorescence/fluorescenceintro.html.

(66) Abramowitz, M., http://microscope.fsu.edu/primer/techniques/fluorescence/fluorointrohome.html.

(67) Grynkiewicz, G.; Poenie, M.; Tsien, R. Y. *J. Biol. Chem.* **1985**, *260*, 3440-3450.

(68) Kurebayashi, N.; Harkins, A. B.; Baylor, S. M. *Biophys. J.* **1993**, *64*, 1934-1960.

(69) Knight, M. R.; Read, N. D.; Campbell, A. K.; Trewavas, A. J. *J. Cell Biol.* **1993**, *121*, 83-90.

(70) Rutter, G. A.; Burnett, P.; Rizzuto, R.; Brini, M.; Murgia, M.; Pozzan, T.; Tavare, J. M.; Denton, R. M. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 5489-5494.

(71) Miller, A. L.; Karplus, E.; Jaffe, L. F. *Methods Cell. Biol.* **1994**, *40*, 305-338.

(72) Wu, Y.; Clusin, W. T. *Am. J. Physiol.* **1997**, *273*, H2161-2169.

(73) Foyouzi-Youssefi, R.; Arnaudeau, S.; Borner, C.; Kelley, W. L.; Tschopp, J.; Lew, D. P.; Demaurex, N.; Krause, K. H. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5723-5728.

(74) Giepmans, B. N.; Adams, S. R.; Ellisman, M. H.; Tsien, R. Y. *Science* **2006**, *312*, 217-224.

(75) Cormack, B. P.; Valdivia, R. H.; Falkow, S. *Gene* **1996**, *173*, 33-38.

(76) Patterson, G. H.; Knobel, S. M.; Sharif, W. D.; Kain, S. R.; Piston, D. W. *Biophys. J.* **1997**, *73*, 2782-2790.

(77) Cubitt, A. B.; Woollenweber, L. A.; Heim, R. *Methods Cell Biol.* **1999**, *58*, 19-30.

(78) Zapata-Hommer, O.; Griesbeck, O. *BMC Biotechnol.* **2003**, *3*, 5.

(79) Rizzo, M. A.; Springer, G. H.; Granada, B.; Piston, D. W. *Nat. Biotechnol.* **2004**, *22*, 445-449.

(80) Nguyen, A. W.; Daugherty, P. S. *Nat. Biotechnol.* **2005**, *23*, 355-360.

(81) Wachter, R. M.; Elsliger, M. A.; Kallio, K.; Hanson, G. T.; Remington, S. J. *Structure* **1998**, *6*, 1267-1277.

(82) Griesbeck, O.; Baird, G. S.; Campbell, R. E.; Zacharias, D. A.; Tsien, R. Y. *J. Biol. Chem.* **2001**, *276*, 29188-29194.

(83) Nagai, T.; Ibata, K.; Park, E. S.; Kubota, M.; Mikoshiba, K.; Miyawaki, A. *Nat. Biotechnol.* **2002**, *20*, 87-90.

(84) Rekas, A.; Alattia, J. R.; Nagai, T.; Miyawaki, A.; Ikura, M. *J Biol Chem* **2002**, *277*, 50573-50578.

(85) Steele, D. *HHMI Bulletin* **2004**, 22-26.

(86) Shkrob, M. A.; Yanushevich, Y. G.; Chudakov, D. M.; Gurskaya, N. G.; Labas, Y. A.; Poponov, S. Y.; Mudrik, N. N.; Lukyanov, S.; Lukyanov, K. A. *Biochem. J.* **2005**, *392*, 649-654.

(87) Wang, D.; Jensen, R. H.; Williams, K. E.; Pallavicini, M. G. *Proteomics* **2004**, *4*, 2175-2183.

(88) Waterman-Storer, C. M.; Desai, A.; Bulinski, J. C.; Salmon, E. D. *Curr. Biol.* **1998**, *8*, 1227-1230.

(89) Servant, G.; Weiner, O. D.; Herzmark, P.; Balla, T.; Sedat, J. W.; Bourne, H. R. *Science* **2000**, *287*, 1037-1040.

(90) Topell, S.; Hennecke, J.; Glockshuber, R. *FEBS Lett.* **1999**, *457*, 283-289.

(91) Jayaraman, S.; Haggie, P.; Wachter, R. M.; Remington, S. J.; Verkman, A. S. *J. Biol. Chem.* **2000**, *275*, 6047-6650.

(92) Akemann, W.; Raj, C. D.; Knopfel, T. *Photochem. Photobiol.* **2001**, *74*, 356-363.

(93) Shimozono, S.; Fukano, T.; Nagai, T.; Kirino, Y.; Mizuno, H.; Miyawaki, A. *Sci STKE* **2002**, *2002*, PL4.

(94) Nagai, T.; Yamada, S.; Tominaga, T.; Ichikawa, M.; Miyawaki, A. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 10554-10559.

(95) Wang, T.; Sakai, Y.; Nakajima, K.; Miyawaki, A.; Ito, K.; Hara, M. *Colloids Surf B Biointerfaces* **2005**, *40*, 183-187.

(96) Belousov, V. V.; Fradkov, A. F.; Lukyanov, K. A.; Staroverov, D. B.; Shakhbazov, K. S.; Terskikh, A. V.; Lukyanov, S. *Nat. Methods* **2006**, *3*, 281-286.

(97) Chiang, J. J.; Li, I.; Truong, K. *Biotechnol. Lett.* **2006**, *28*, 471-475.

(98) Hicks, B. W.; Humana Press, 2002.

(99) Goldenberg, D. P.; Creighton, T. E. *J. Mol. Biol.* **1983**, *165*, 407-413.

(100) Romoser, V. A.; Hinkle, P. M.; Persechini, A. *J. Biol. Chem.* **1997**, *272*, 13270-13274.

(101) Heim, R.; Tsien, R. Y. *Curr. Biol.* **1996**, *6*, 178-182.

(102) Demarco, I. A.; Periasamy, A.; Booker, C. F.; Day, R. N. *Nat. Methods.* **2006**, *3*, 519-524.

(103) Wang, J. W.; Wong, A. M.; Flores, J.; Vosshall, L. B.; Axel, R. *Cell* **2003**, *112*, 271-282.

(104) Yu, D.; Baird, G. S.; Tsien, R. Y.; Davis, R. L. *J. Neurosci.* **2003**, *23*, 64-72.

(105) Richmond, T. A.; Takahashi, T. T.; Shimkhada, R.; Bernsdorf, J. *Biochem. Biophys. Res. Commun.* **2000**, *268*, 462-465.

(106) Rapizzi, E.; Pinton, P.; Szabadkai, G.; Wieckowski, M. R.; Vandecasteele, G.; Baird, G.; Tuft, R. A.; Fogarty, K. E.; Rizzuto, R. *J. Cell. Biol.* **2002**, *159*, 613-624.

# CHAPTER 2

# SELECTION OF MCHERRY VARIANTS WITH A PEPTIDE INSERTION

## 2.1 INTRODUCTION

The red fluorescent protein, drFP583 (commercial name DsRed), cloned from *Discosoma* coral [1], has become very useful in biotechnology and cell biology because it extends the spectrum of FP colors to red wavelengths and provides a distinct label for multicolor tracking of fusion proteins together with GFP from the *Aequorea* jellyfish [2]. However, a slow, incomplete maturation and obligate tetramerization [3] impede its use in many applications. In 2002, a true monomeric DsRed [4], named mRFP1, was generated by multiple steps of random and site-directed mutagenesis. To satisfy the application requirements and extend the color palette, a "fruit" series of fluorescent proteins have been made. Among the true monomeric "fruit" series of fluorescent proteins, mCherry offers the longer wavelengths (3 nm red shift to mRFP1), the highest photostability, the fastest maturation and excellent pH resistance. It has higher excitation although its quantum efficiency is slightly lower.

Domain insertions have greatly contributed to protein evolution, although it is not easy to predict the effects of amino acid insertions on the structure and function of proteins. Several studies indicated that sites tolerant to a few amino acid insertions are observed almost exclusively in loop regions [5-8] and rarely within α-helices [9, 10]. The screening method combined with the insertional gene fusion technique [11, 12] has been adopted to insert a foreign epitope sequence into a surface loop of enzymes. The first insertion in GFP was a BLIP (β–lactamase inhibitory protein) sensor [13] in which Bla (β-lactamase) was

genetically inserted into GFP between the Gln-172 and Asp-173 sites, which is located at a solvent-exposed loop between β-strands and tolerates 6–20 amino acid insertions without serious disturbance of the GFP structure and function. In addition, a six-residue peptide (FKTRHN) was inserted into CFP at position 145 without destroying its fluorescence during some semi-random mutagenesis experiments [14]. Similarly, it might be expected that the mCherry protein should tolerate some site insertions and stay fluorescent. In the present study, 15 base pairs were randomly inserted into mCherry by Mutation Generation System™ Kit (Finnzymes) to build a library to search the promising insertion sites.

The Mutation Generation System™ Kit (MGS™ Kit) is designed for rapid construction of insertion mutation libraries for any kind of DNA clones. In the system a pool of 15 bp random insertion mutants is generated by a simple *in vitro* transposition reaction catalyzed by a single purified enzyme, MuA Transposase (Figure 2.1 A I). Four monomers of the MuA Transposase protein are bound to the R1 and R2 sites at the end of the Entranceposon (Figure 2.1 A II), an artificial transposon, to assemble the *in vitro* Mu transposition complex (Figure 2.1 C) with sticky end. The MuA Transposase makes a 5 bp staggered cut in the target DNA (Figure 2.1 B) and at the same time the 3' end of the Entranceposon is covalently joined to phosphates in the target site to form transposition complex (Figures 2.1 D and E). The machinery of the bacteriophage Mu randomly selects a 5 bp target site for transposition (Figure 2.2 A) and replicates the 5 bps by transposing repeatedly inside the host genome and inserts Entranceposon into the target plasmid in random sites (Figure 2.2 B). The insertion clones are digested with the

rare-cutting restriction enzyme Not I to remove the body of the Entranceposon (Figure 2.1 E). Closure of the Not I digested clones by self-ligation results in a 15 bp insertion in the target DNA (Figure 2.1 F), of which 5 bp are duplicated from the target DNA and 10 bp are from the Entranceposon (Figure 2.2 C).

In this thesis work, the Mutation Generation System (MGS™, Finnzymes) was used to insert 15 bp sequences randomly into the gene encoding mCherry to create a linker scanning library of several thousand members. Most insertions had deleterious effects on mCherry folding and fluorescence, which is consistent with reports for other fluorescent proteins [13]. However, through library screening, six locations within mCherry were identified that maintained red fluorescence despite the insertion. These locations are "permissive sites" that are likely to tolerate the N- and C-termini of a circular permuted variant (see Chapter 3).

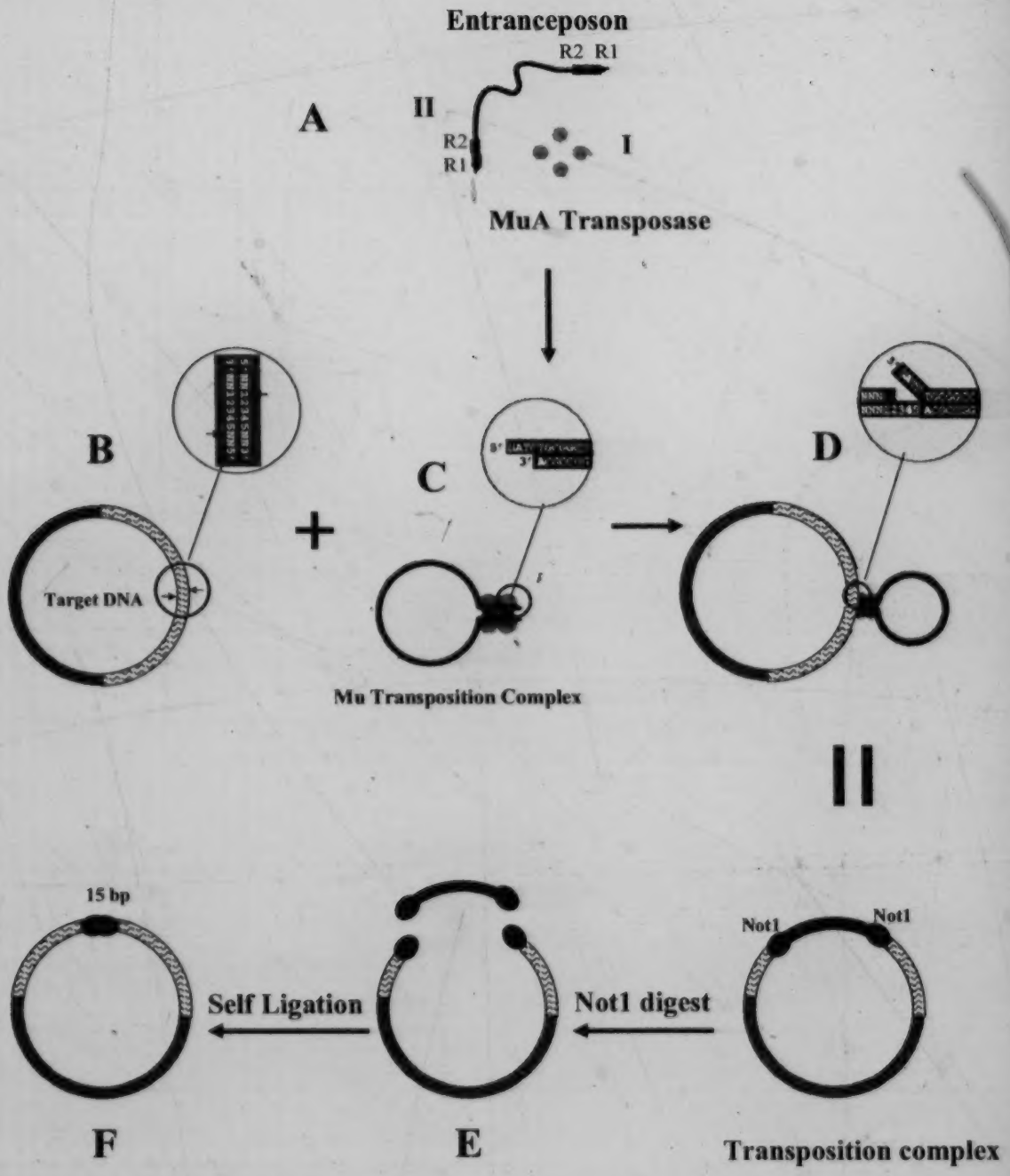



**Figure 2.1** Flow chart of the MGS protocol of *in vitro* transposition reaction.

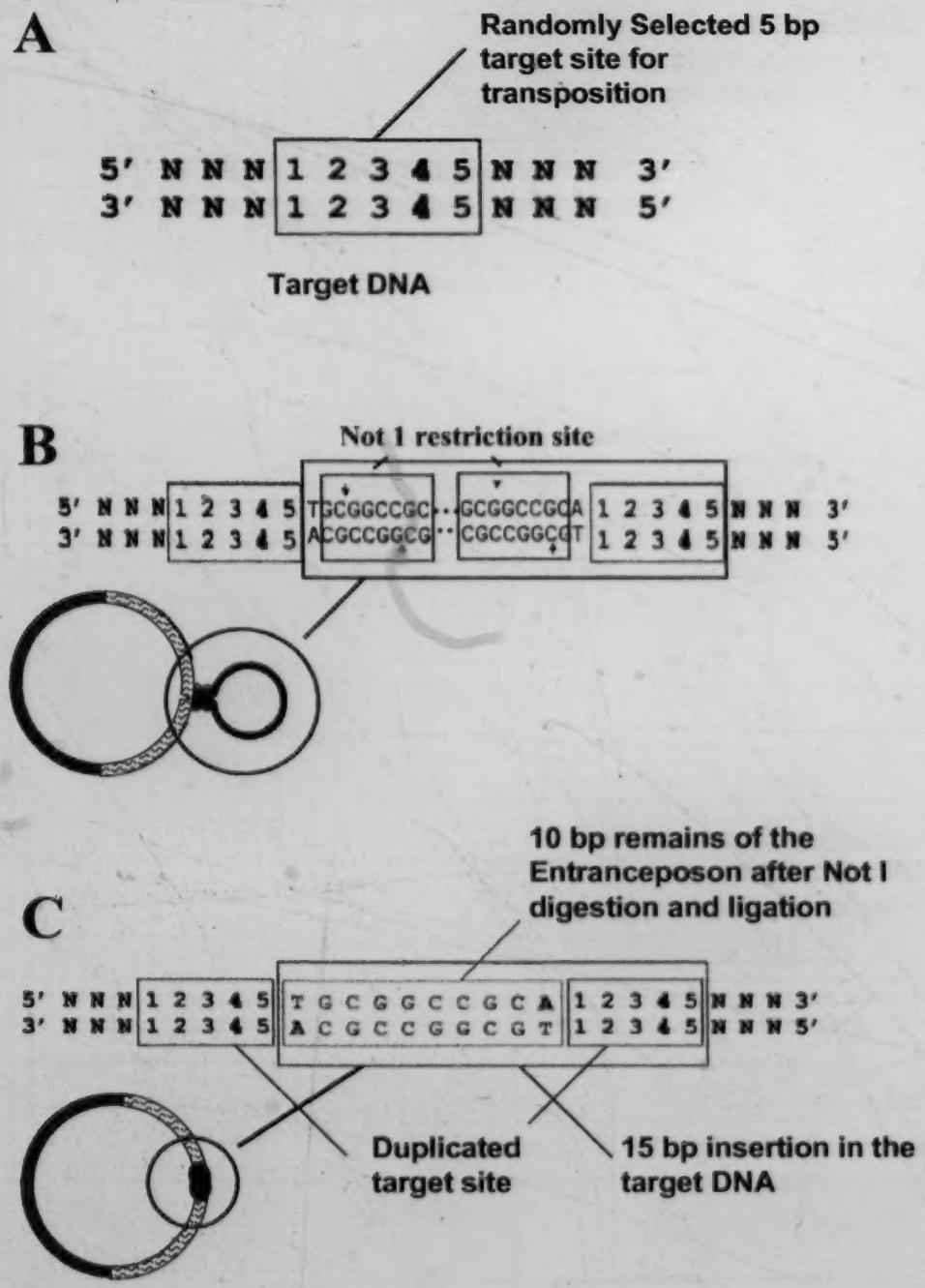

A
Randomly Selected 5 bp target site for transposition
5′ N N N 1 2 3 4 5 N N N 3′
3′ N N N 1 2 3 4 5 N N N 5′
Target DNA
B
Not I restriction site
5′ N N N 1 2 3 4 5 TGCGGCCGC··GCGGCCGCA 1 2 3 4 5 N N N 3′
3′ N N N 1 2 3 4 5 ACGCCGGCG··CGCCGGCGT 1 2 3 4 5 N N N 5′
C
10 bp remains of the Entranceposon after Not I digestion and ligation
5′ N N N 1 2 3 4 5 T G C G G C C G C A 1 2 3 4 5 N N N 3′
3′ N N N 1 2 3 4 5 A C G C C G G C G T 1 2 3 4 5 N N N 5′
Duplicated target site
15 bp insertion in the target DNA

**Figure 2.2** Structure of the 15 bp insertion generated by the MGS. (A) Target DNA; (B) Target DNA with Entranxeposon insertion (transposition complex); (C) Target DNA with 15 bp insertion.

## 2.2 EXPERIMENTAL

### 2.2.1 GENERAL METHODS AND MATERIALS

A synthetic gene of mCherry1 was a gift from Dr. Tsien's laboratory, Department of Pharmacology, and Department of Chemistry & Biochemistry, University of California, San Diego, USA. All synthetic DNA oligonucleotides for cloning and construction of subsequent libraries were purchased from Sigma-Genosys Canada or Integrated DNA Technologies. Mutation Generation System™ Kit (MGS™ Kit) was purchased from Finnzymes. Pfu polymerase (Fermentas) was used for PCR amplifications in the buffer supplied by the manufacturer. Other modifying enzymes and restriction enzymes were purchased from either Invitrogen or New England Biolabs. PCR products and products of restriction digests were routinely purified using the QIAquick PCR purification kit (Qiagen) according to the manufacturer's protocols. All the sequences were confirmed by dye terminator cycle sequencing using the DYEnamic ET kit (Amersham Biosciences). Sequencing reactions were analyzed at the University of Alberta Molecular Biology Service Unit.

### 2.2.2 GENE CONSTRUCTION

The gene of mCherry was amplified by polymerase chain reaction (PCR) with a sense primer containing the Xba1-Kpn1 sites and encoding a peptide linker GGS: gg tctaga ggtacc *ggcggctcc* atggtgagcaagggcgag (Xba1 and Kpn1 sites are underlined, and GGS is in italics) and a reverse primer containing the EcoR1-Kpn1 sites and encoding a peptide linker GGT: g gaattc *ggtacc gcc*

cttgtacagctcgtccatgcc (EcoR1 and Kpn1 sites underlined, and GGT in italics). The PCR product was digested with Xba1 and Xho1, and ligated with T4 DNA ligase (Invitrogen) into similarly digested pUC18 vector (Invitrogen). The ligated product was used to transform *E. coli* DH10B (Invitrogen) by electroporation. Appropriate dilutions of transformed bacteria were plated on LB/agar containing ampicillin (0.1 mg/mL) and after overnight incubation at 37 °C, single colonies were picked and used to inoculate 4 mL of LB media containing ampicillin (0.1 mg/mL). Cultures were grown with shaking overnight at 37 °C before plasmids were isolated using the QIAprep spin miniprep kit according to the manufacturer's protocols (Qiagen).

### 2.2.3 RANDOM 15 BP INSERTION MUTATION LIBRARY OF MCHERRY CONSTRUCTION

The Mutation Generation System™ Kit (MGS™ Kit, Finnzymes) was used to construct the insertion mutation library of mCherry variants.

According to the manufacturer's protocols, *in vitro* transposition reaction needs an optimal amount of the target DNA depending on the size of the plasmid calculated as: Amount of the target DNA (ng) per reaction = Size of the target plasmid (kb) × 40 ng. So, mCherry DNA plasmid consists of a 0.7 kb insert cloned into a 2.7 kb vector. The size of the mCherry plasmid is 0.7 kb + 2.7 kb = 3.4 kb, and the optimal amount per reaction is: 3.4 × 40 ng = 136 ng. The concentration of mCherry plasmid was 125 ng/μL, so a minimum of 136 ng, equal to 1.1 μL, of the target DNA per transposition reaction, was used. The *in vitro* transposition reaction was set up following the manufacturer's protocols and

Entranceposon (M1-Kan$^R$) was used here. The reaction mixture was diluted ten-fold in deionized water and transformed into electrocompetent *E. coli* strain DH10B (Invitrogen) and plated out on 5 Luria-Bertani (LB)/agar plates supplemented with ampicillin (0.1 mg/mL) and Kanamycin (10 μg/mL). Plates were incubated for 14 h at 37 °C.

The transformants were scraped from the plates and pooled bacteria and plasmid DNA were prepared using the QIAquick miniprep kit (Qiagen) according to the manufacturer's protocols. The plasmid was digested with Xba1 and EcoR1. The resulting fragments were separated by standard agarose gel electrophoresis (Figure 2.5 A). The DNA of interest was extracted from the gel by using a QIAquick gel extraction kit (Qiagen) according to the manufacturer's protocols and ligated into similarly digested pBAD/His B vector (Invitrogen). Electrocompetent *E. coli* strain DH10B (Invitrogen) was transformed and plated out on 5 Luria-Bertani (LB)/agar plates supplemented with ampicillin (0.1 mg/mL) and Kanamycin (10 μg/mL). Plates were incubated for 14 h at 37 °C.

The plasmid DNA was prepared as described above and digested with Not1. The resulting fragments were separated by standard agarose gel electrophoresis (Figure 2.5 B). The fragment of mCherry with the 15 bp gene insertion was extracted from the gel as previously described and self-ligated with T4 ligase. Electrocompetent *E. coli* strain DH10B (Invitrogen) was transformed and plated on Luria-Bertani (LB)/agar plates supplemented with ampicillin (0.1 mg/mL). Plates were incubated for 14 h for 37 °C. The plasmid DNA was prepared as described above to provide the mCherry insertion plasmid library.

### 2.2.4 LIBRARY SCREENING

The mCherry insertion plasmid library was transformed into electrocompetent *E. coli* strain DH10B (Invitrogen) and plated on Luria-Bertani (LB)/agar plates supplemented with ampicillin (0.1 mg/mL) and L-arabinose (0.02%). Plates were incubated for 14 h at 37 °C prior to screening.

The system for imaging the fluorescence of bacterial colonies grown on 10 cm Petri dishes is a custom built device that has been described in detail elsewhere [15, 16]. In brief, the light from a 175 W xenon arc lamp (Sutter) is passed through a 560/40 nm bandpass filter (Chroma) and into a bifurcated fiber-optic bundle (Newport). Light exiting the fiber-optic bundle illuminates a 10 cm diameter area with an irradiance of approximately 0.04 $mW/cm^2$. For all screening up to the identification of the insertion mCherrry, colony fluorescence on illuminated plates was viewed through a pair of red wraparound goggles that block light of less than ~580 nm (Lightning Powder Company, Inc.). When viewed through these goggles, colonies fluorescing at 580 nm or longer wavelength show their red fluorescence and can be distinguished. One hundred bright or dim red fluorescent colonies were picked, retransformed and the plasmid DNA (numbered from #1 to #100) was prepared.

### 2.2.5 15 BP INSERTION MAPPING BY COLONY PCR OR DIGEST

For reliable mapping of the 15 bp insertion in the target DNA, two separate 20 µL PCR reactions per insertion clone were performed (Figure 2.7) following the manufacturer's protocols. However, it was found that it was easier

to map the position of the 15 bp insertion using restriction digestion. The digestion protocol was found to be more reliable than PCR, the restriction enzyme is cheaper than Tap DNA polymerase, and the digest is much faster than PCR. For the digest mapping, three separate 10μL digest reactions were performed per insertion plasmid: the first one was digested with Not1 and Xba1, the second one was digested with Not1 and EcoR1, the other was digested with Xba1 and EcoR1 (Figure 2.6). The digested products were analyzed by standard agarose gel electrophoresis. The length of a PCR product equals the distance between the 15 bp insertion site and the restriction site (Xba1 or EcoR1). The mapping of insertions in mCherry was obtained by comparing the lengths of the products. Both the PCR-based method and the restriction digest-based method were used to identify promising plasmids with unique insertion sites. The promising plasmids (#1, #2, #19, #32, #47 and #67) were confirmed by sequence analysis.

### 2.2.6 PROTEIN PURIFICATION AND CHARACTERIZATION

To prepare proteins in sufficient quantity for characterization, *E. coli* strains DH10B or LMG194 were transformed with the pBAD/His B expression vector containing the gene of interest. A single colony was used to inoculate a 4 mL culture that was allowed to grow overnight (37 °C, 225 rpm) before being diluted into 1 L of LB media containing ampicillin (0.1 mg/mL) and arabinose (0.2%). The culture was allowed to grow for 12 h before cells were harvested by centrifugation and lysed by French Press. Proteins were purified by Ni-NTA chromatography (Amersham). In Ni-NTA clolum, $Ni^{2+}$ is attached to NTA resin

and the immobilized $Ni^{2+}$ can strongly bind the histidine residues in the DNA plasmid.

### 2.2.7 SPECTROSCOPY OF THE INSERTION PROTEIN AND THE FLUORESCENCE INTENSITIES OF THE COLONIES

Absorption spectra were recorded on a DU-800 UV-visible spectrophotometer (Beckman). Quantum yields (QY) for insertion mCherry variants were measured using mCherry in 10 mM NaOH as the reference standard [17]. Apparent extinction coefficients (EC') were measured by UV-visible absorbance spectroscopy on purified proteins. To determine the ensemble EC of the total protein (inclusive of polypeptide chains that did not form a red chromophore), the intensity of the 587 nm absorbance peak for each protein was compared to that of a solution of mCherry with matched absorbance at 280 nm. The EC' was calculated by multiplying the ratio of absorbance intensities at 587 nm by the EC of mCherry (72,000 $M^{-1}cm^{-1}$) [18]. To determine the intrinsic EC for only those polypeptide chains that did form a red chromophore, the alkali denaturation method was employed [18, 19]. All emission spectra were acquired on a QuantaMaster spectrofluorometer (Photon Technology International) and have been corrected for the instrument response.

To identify the fluorescence intensity (brightness) of an insertion mCherry, colony fluorescence was digitally imaged with a Retiga 1300i 12-bit CCD camera (QImaging) fitted with a filter wheel (Sutter) that contains both a a 630/60 nm bandpass filter. Through the software of Image Pro Plus (Media Cybernetics),

images were acquired and the fluorescence intensities of all colonies were individually integrated.

## 2.3 RESULTS AND DISCUSSION

### 2.3.1 CONSTRUCTION OF MCHERRY IN PUC18 PLASMID

mCherry gene (Figure 2.3, Line 4) was cloned into pUC18 (Figure 2.3, Line 2), because pUC18 has the properties of small size (2,686 bp in length) and high copy number *Escherichia coli* (*E. coli*) plasmids. The high copy number of pUC plasmids is a result of the lack of the *rop* gene and a single point mutation in *rep* of pMB1. In pUC18, mCherry (Figure 2.3, Line 5) contains XbaI, EcoRI, KpnI sites, and encodes GGS and TGG peptides (GGS-mCherry-GGT).

**Figure 2.3** DNA agarose gel. All samples were subjected to electrophoresis in a 1% agarose gel in a buffer containing 100 mM Tris-HCl, 120 mM sodium borate (pH 8.4), 0.1 mM EDTA, and ethidium bromide (0.5 mg/ml). Line L: 1 Kb plus ladder (Invitrogen). Line 1: pUC18 contains CFP. Line 2: pUC 18 vector which is from the digested fragment of pUC18 containing CFP digested with Xba1 and EcoR1. Line 3: pBAD contains mCherry. Line 4: mCherry gene with Xba1, Kpn1 and EcoR2 restriction site. Line 5: the ligation product of ligating pUC 18 (Line 2) and (Line 4). Line 5: digest of the ligation product with XbaI and EcoRI. Line 6: digest of the ligation product with KpnI.

### 2.3.2. CONSTRUCTION OF 15 BP INSERTION MCHERRRY VARIANTS

Because the transposition reaction of the MGS has no target sequence preference, the Entranceposon insertions occur at random locations in the whole mCherry plasmid. In other words, the insertion could occur in the vector backbone (pUC18) as well as in the gene of interest (mCherry) (shown in Figure 2.4). However, only the Entranceposon insertions occurring in the region of the mCherry gene are useful. Therefore, the separation of the different insertions by standard agarose gel electrophoresis was necessary. The insertion plasmids were digested with Xba1 and EcoR1 and the gel electrophoresis patterns are shown in Figure 2.5 A. After digest, the fragments of plasmid with the insertion within the vectors (Figures 2.4 E, F and G) should be 3.8 kb (vector 2.7 kb + Entranceposon M1-Kan$^R$ 1.1 kb) and 0.7 kb (mCherry gene) in length; the fragments of plasmid with the insertion within the mCherry gene (Figures 2.4 A, B and C) should be: 2.7 kb (vector) and 1.8 kb (mCherry gene 0.7 kb + Entranceposon M1-Kan$^R$ 1.1 kb) in length (Figure 2.6); and the fragments of plasmid with the insertion within EcoR 1 site or Xba 1 site (Figures 2.4 D and H), should be 4.5 kb (vector 2.7 kb + mCherry gene 0.7 kb + Entranceposon M1-Kan$^R$ 1.1 kb) in length. The band of 1.8 kb is the band of interest, which was cut and ligated into fresh pBAD/His B vector. After digest with Not1 (Gel electrophoresis patterns shown in Figure 2.5 B), the Entranceposon M1-Kan$^R$ gene (1.1 kb) was removed, and the linear gene of mCherry with pBAD/His B (4.8 kb) was extracted to self-ligate to build the

gene library of 15 bp insertion mCherry variants (the principle shown in Figure 2.1 and 2.2).

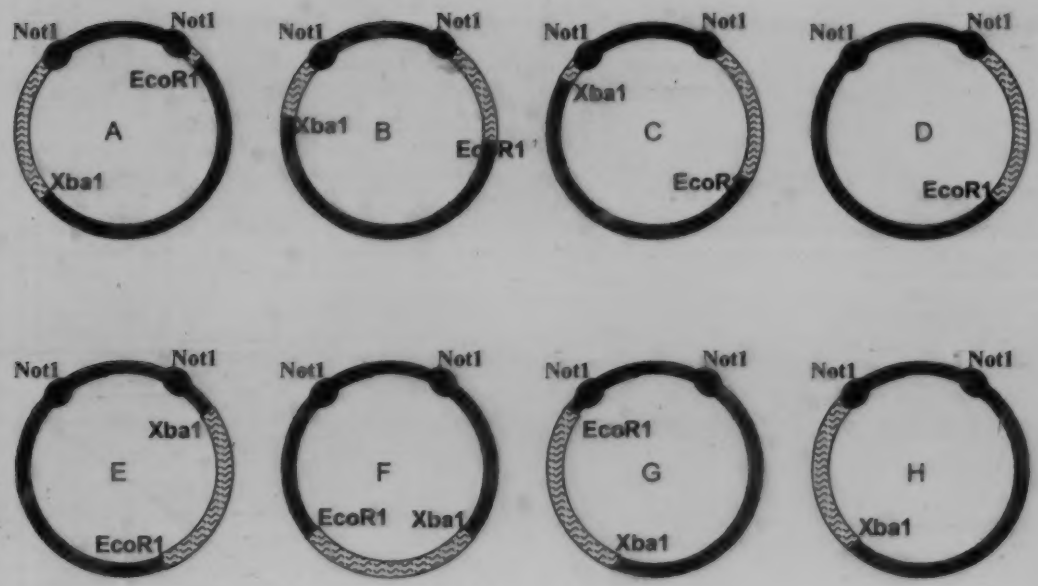



**Figure 2.4** Pool of target mCherry clones carrying an Entranceposon insertion at random locations. is mCherry gene, is pUC18 vector. A, B and C, insertions occur within mCherry gene. D and H, insertions occur between mCherry and pUC18 vector. E, F and G, insertions occur within pUC18 vector.

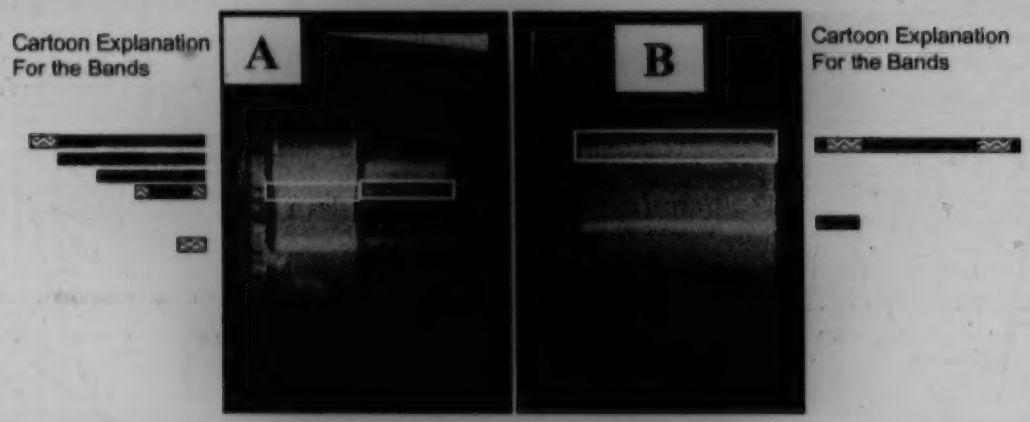



**Figure 2.5** Gel electrophoresis patterns of Entranceposon insertion mCherry digest. All samples were subjected to electrophoresis in a 1% agarose gel in a buffer containing 100 mM Tris-HCl, 120 mM sodium borate (pH 8.4), 0.1 mM EDTA, and ethidium bromide (0.5 mg/mL). Lane 1 of each gel: 1 kb plus DNA ladder (Invitrogen in A and Fermentas in B). A shows the pattern for the fragments of and Entranceposon insertion mCherry variants digested by EcoR1 and Xba1. B shows the fragments of Entranceposon insertion mCherry digested by Not1. ▒ is mCherry gene, ▬ is pUC18 vector, ▬ is the isertion gene, Entranceposon. ▭ Shows the band of interest which was cut from the gel.

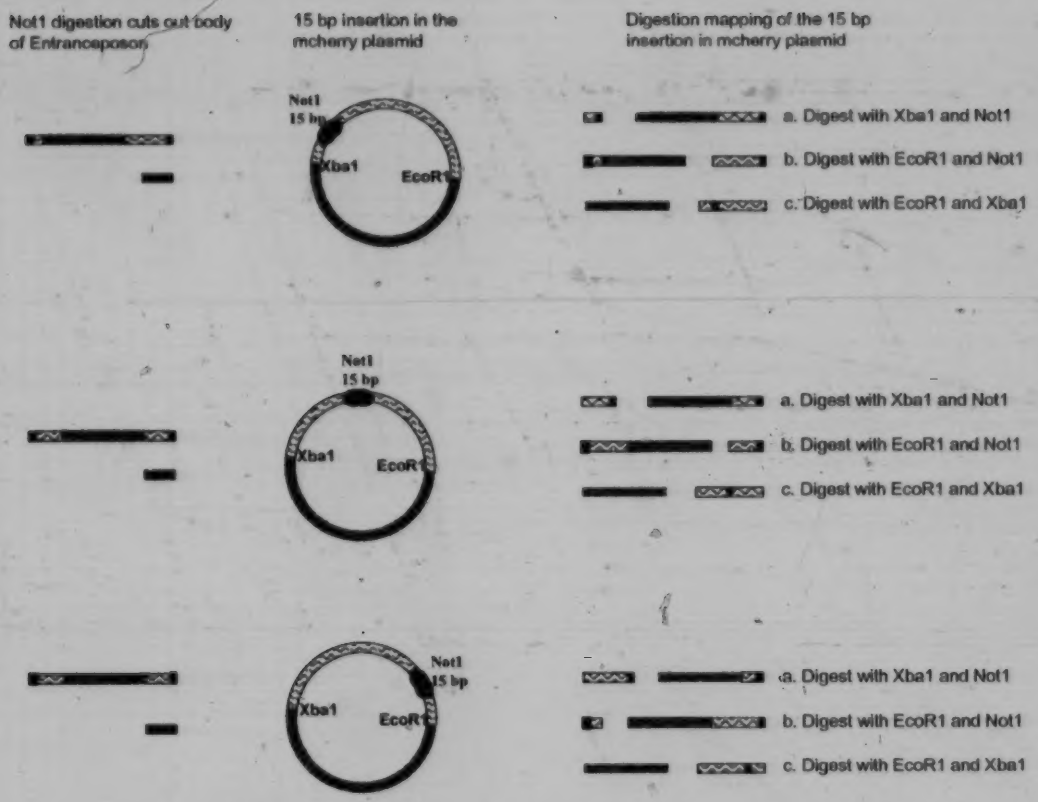

Not1 digestion cuts out body of Entranceposon
15 bp insertion in the mcherry plasmid
Digestion mapping of the 15 bp insertion in mcherry plasmid
Not1
15 bp
Xba1
EcoR1
a. Digest with Xba1 and Not1
b. Digest with EcoR1 and Not1
c. Digest with EcoR1 and Xba1

**Figure 2.6** Cartoons of 1) Not1 digest cuts out the body of Entranceposon; 2) 15 bp insertion in the mCherry plasmid; and 3) Digest mapping of the 15 bp insertion in mCherry plasmid.

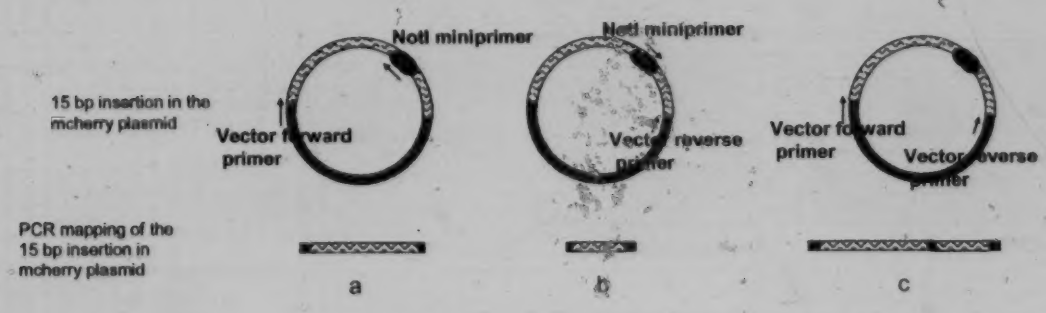
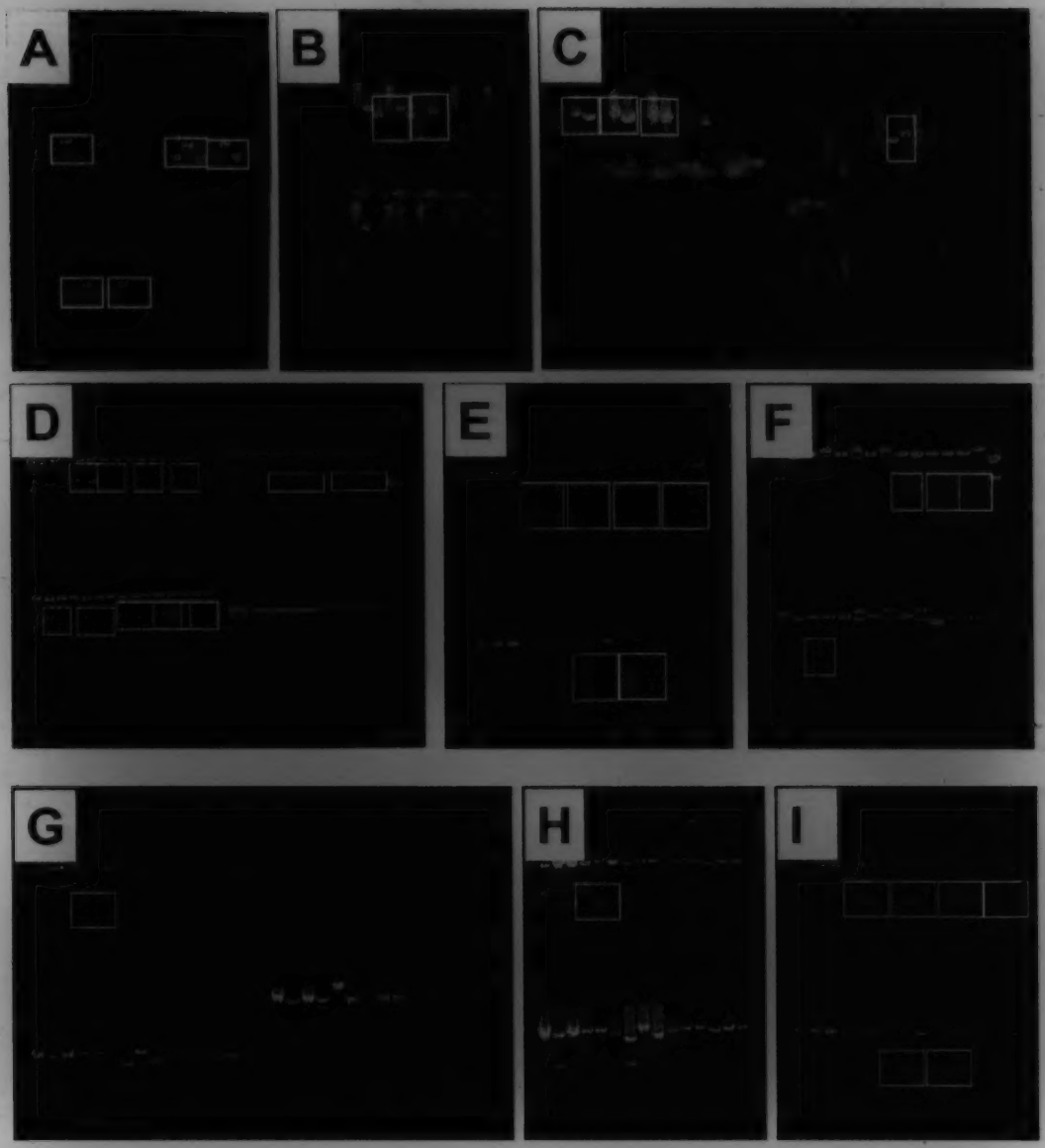



**Figure 2.7** PCR mapping of the 15 bp insertion in mCherry plasmid.

Through the library screen, we determined that greater than 80% of colonies of the insertion mCherry had deleterious effects on mCherry folding and/or fluorescence, as previously reported for other fluorescent proteins [13]. One hundred red fluorescent colonies were selected and prepared with the plasmids numbered from #01 to #100. In order to get the mapping of the 15 bp insertion mCherry plasmid, three (Figure 2.7 a, b and c) or two mapping (Figure 2.7 a and b) PCRs were performed. By comparing the length of the PCR products, the site of the 15 bp could be figured out. Digests also can map the 15 bp insertion by three (Figure 2.6 a, b and c) or two digests (Figure 2.6 a and b). The gel electrophoresis patterns of the PCR products and digest fragments are shown in Figure 2.8. After the mapping, 42 plasmids (boxed or circled in figure 2.8) were selected. After sequence confirmed, six plasmids were chosen, numbered #1, #2, #19, #32, #47 and #67, in which AAAQK was inserted between K138 and K139, MRPQQ was inserted between Q137 and K138, DAAAH was inserted between H25 and E26, AAASV was inserted between V22 and N23, CGRTY was inserted between Y193 and N194, and NAAAA was inserted between A184 and K185 within mCherry, respectively. The sequences of these insertion mCherry mutants compared to that of mCherry is shown in Figure 2.9. As expected, all the six positions are in loops between β-stranded sheets (Figures 2.10 and 2.11).


A
B
C
D
E
F
G
H
I

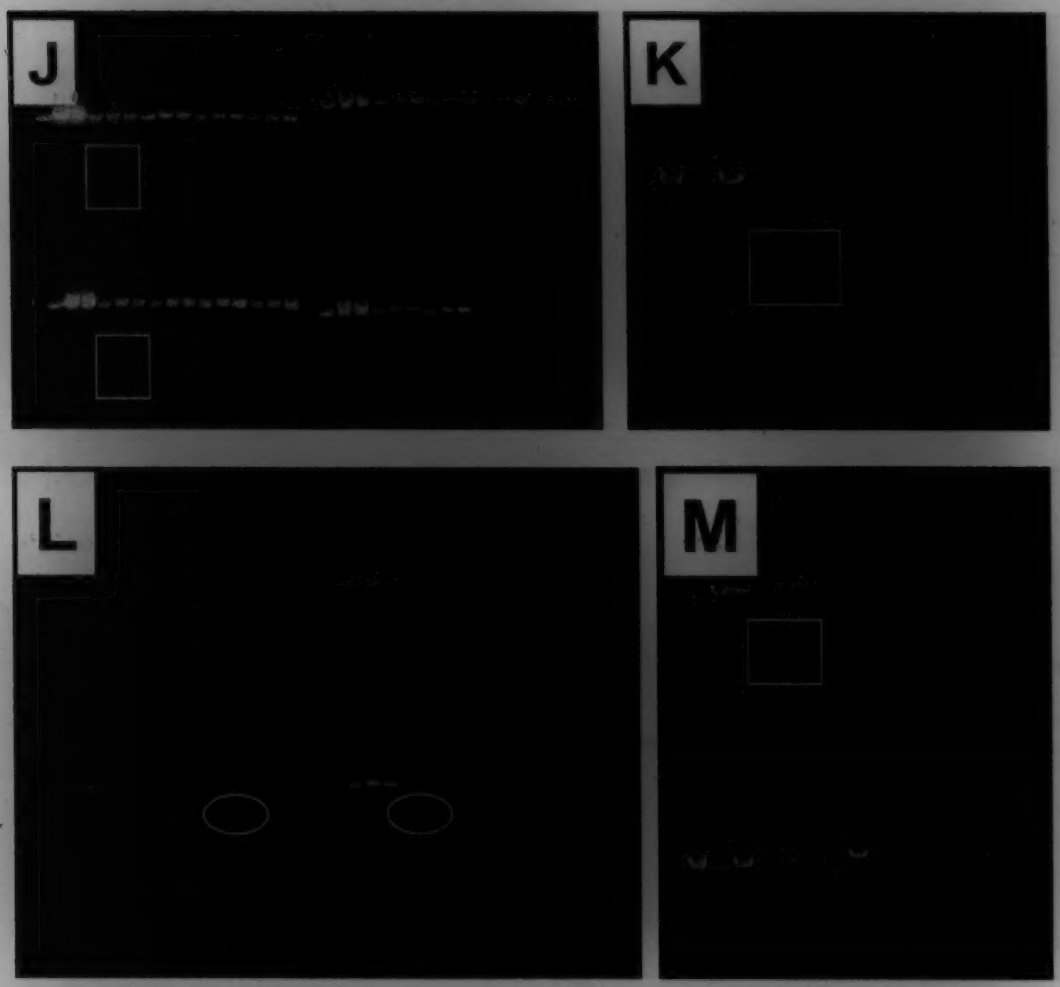
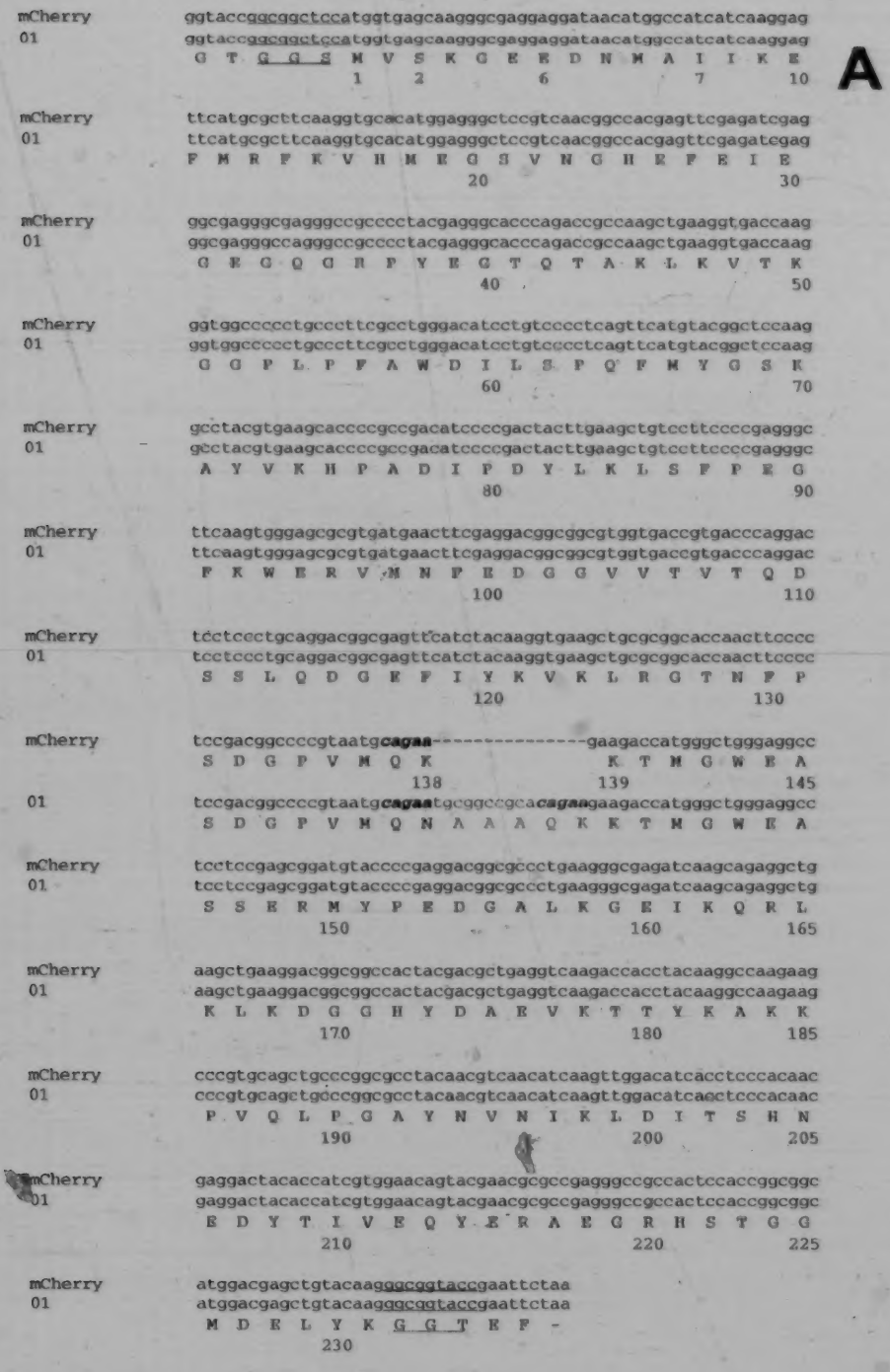
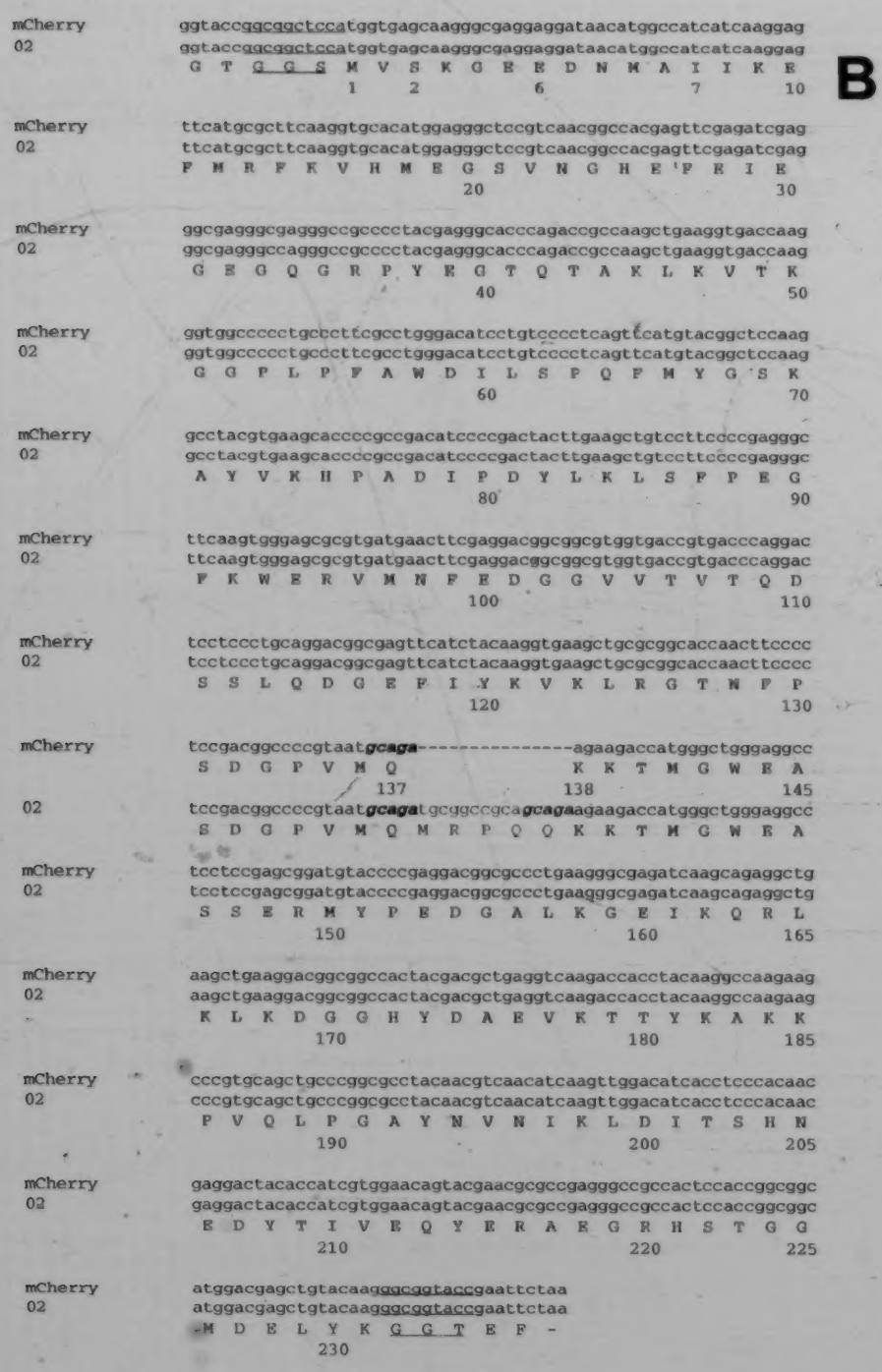
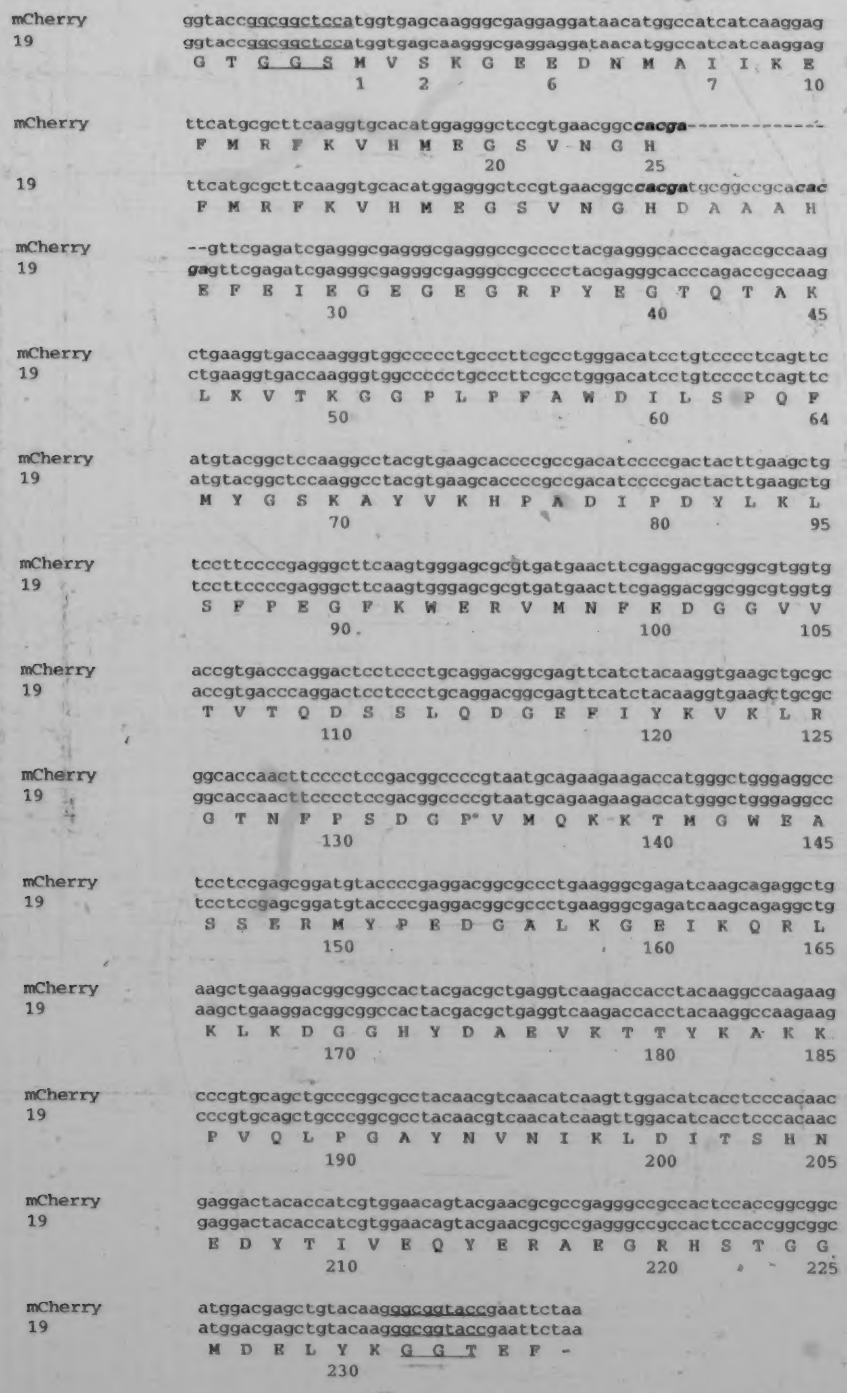


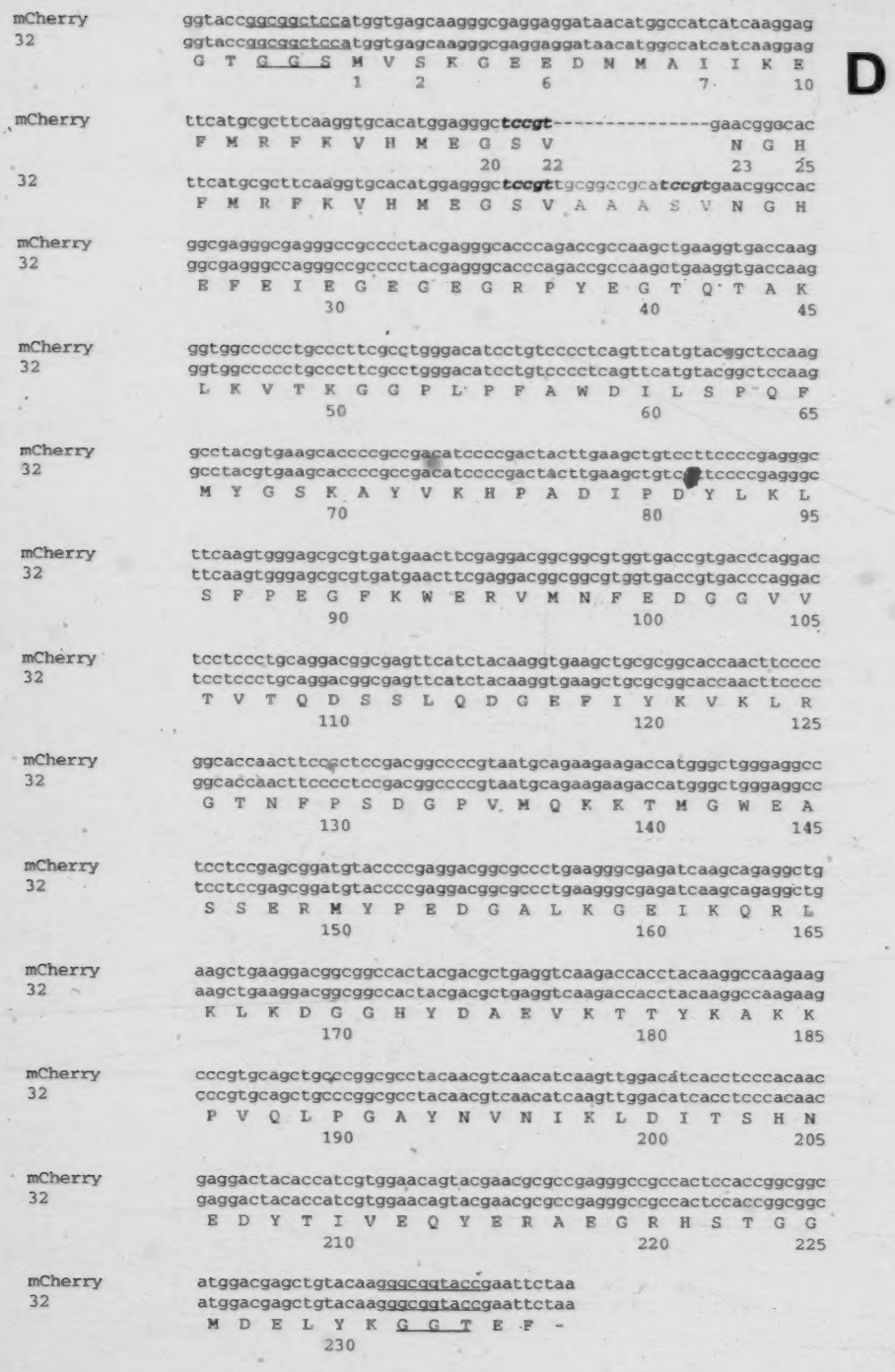

**Figure 2.8** Gel electrophoresis patterns of PCR products or digest fragments with mapping of the 15 bp insertion mCherry. All samples were subjected to electrophoresis in a 1% agarose gel in a buffer containing 100 mM Tris-HCl, 120 mM sodium borate (pH 8.4), 0.1 mM EDTA, and ethidium bromide (0.5 mg/mL). A-C are mapping by PCR. A and B, three PCR (a, c and b (Figure 2.7) in order) were performed. C, two PCR (a and b) were performed. D-M are mapping by digest. D and F, two digests (a and b (Figure 2.6) in order) were performed. Others, three digests (c, a and b in order) were performed. G-M, the first three columns were the #01, which were confirmed by sequence, as reference.

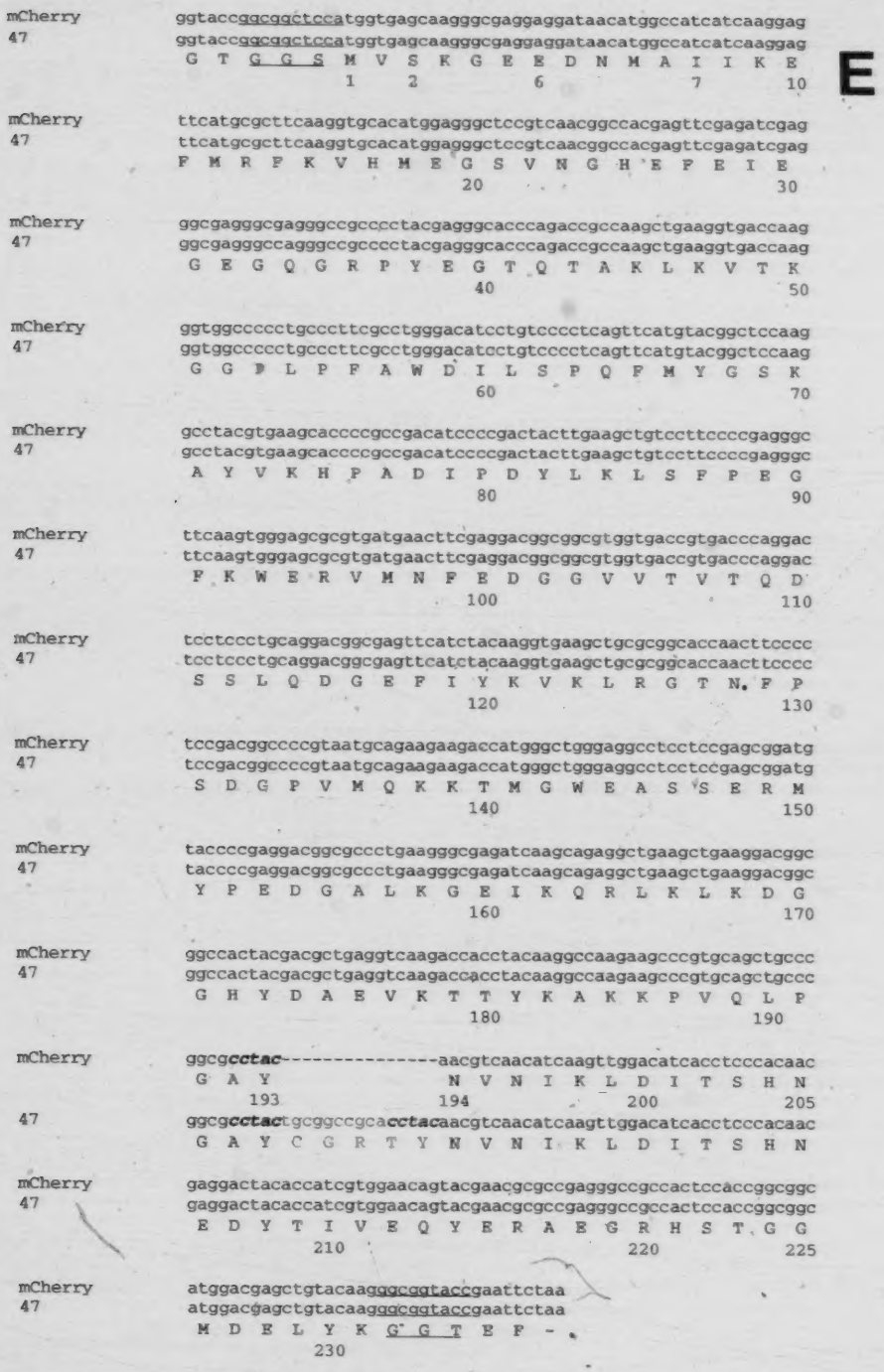

mCherry
01
A
mCherry
01
mCherry
01
mCherry
01
mCherry
01
mCherry
01
mCherry
01
mCherry
01
mCherry
01
mCherry
01
mCherry
01
mCherry
01


mCherry
02
B
1 2 6 7 10
20 30
40 50
60 70
80 90
100 110
120 130
137 138 145
150 160 165
170 180 185
190 200 205
210 220 225
230


mCherry
19
G T G G S M V S K G E E D N M A I I K E
1 2 6 7 10
mCherry
F M R F K V H M E G S V N G H
20 25
19
F M R F K V H M E G S V N G H D A A A H
mCherry
19
E F E I E G E G E G R P Y E G T Q T A K
30 40 45
mCherry
19
L K V T K G G P L P F A W D I L S P Q F
50 60 64
mCherry
19
M Y G S K A Y V K H P A D I P D Y L K L
70 80 95
mCherry
19
S F P E G F K W E R V M N F E D G G V V
90 100 105
mCherry
19
T V T Q D S S L Q D G E F I Y K V K L R
110 120 125
mCherry
19
G T N F P S D G P V M Q K K T M G W E A
130 140 145
mCherry
19
S S E R M Y P E D G A L K G E I K Q R L
150 160 165
mCherry
19
K L K D G G H Y D A E V K T T Y K A K K
170 180 185
mCherry
19
P V Q L P G A Y N V N I K L D I T S H N
190 200 205
mCherry
19
E D Y T I V E Q Y E R A E G R H S T G G
210 220 225
mCherry
19
M D E L Y K G G T E F -
230

C


mCherry
32
ggtaccggcggctccatggtgagcaagggcgaggaggataacatggccatcatcaaggag
ggtaccggcggctccatggtgagcaagggcgaggaggataacatggccatcatcaaggag
G T G G S M V S K G E E D N M A I I K E
1 2 6 7 10
D
mCherry
ttcatgcgcttcaaggtgcacatggagggctccgt----------gaacggccac
F M R F K V H M E G S V N G H
20 22 23 25
32
ttcatgcgcttcaaggtgcacatggagggctccgttgcggccgcatccgtgaacggccac
F M R F K V H M E G S V A A A S V N G H
mCherry
ggcgagggcgagggccgcccctacgagggcacccagaccgccaagctgaaggtgaccaag
32
ggcgagggccagggccgcccctacgagggcacccagaccgccaagctgaaggtgaccaag
E F E I E G E G E G R P Y E G T Q T A K
30 40 45
mCherry
ggtggcccccctgcccttcgcctgggacatcctgtcccctcagttcatgtacggctccaag
32
ggtggcccccctgcccttcgcctgggacatcctgtcccctcagttcatgtacggctccaag
L K V T K G G P L P F A W D I L S P Q F
50 60 65
mCherry
gcctacgtgaagcaccccgccgacatccccgactacttgaagctgtccttccccgagggc
32
gcctacgtgaagcaccccgccgacatccccgactacttgaagctgtccttccccgagggc
M Y G S K A Y V K H P A D I P D Y L K L
70 80 95
mCherry
ttcaagtgggagcgcgtgatgaacttcgaggacggcggcgtggtgaccgtgacccaggac
32
ttcaagtgggagcgcgtgatgaacttcgaggacggcggcgtggtgaccgtgacccaggac
S F P E G F K W E R V M N F E D G G V V
90 100 105
mCherry
tcctccctgcaggacggcgagttcatctacaaggtgaagctgcgcggcaccaacttcccc
32
tcctccctgcaggacggcgagttcatctacaaggtgaagctgcgcggcaccaacttcccc
T V T Q D S S L Q D G E F I Y K V K L R
110 120 125
mCherry
ggcaccaacttcccctccgacggccccgtaatgcagaagaagaccatgggctgggaggcc
32
ggcaccaacttcccctccgacggccccgtaatgcagaagaagaccatgggctgggaggcc
G T N F P S D G P V M Q K K T M G W E A
130 140 145
mCherry
tcctccgagcggatgtaccccgaggacggcgccctgaagggcgagatcaagcagaggctg
32
tcctccgagcggatgtaccccgaggacggcgccctgaagggcgagatcaagcagaggctg
S S E R M Y P E D G A L K G E I K Q R L
150 160 165
mCherry
aagctgaaggacggcggccactacgacgctgaggtcaagaccacctacaaggccaagaag
32
aagctgaaggacggcggccactacgacgctgaggtcaagaccacctacaaggccaagaag
K L K D G G H Y D A E V K T T Y K A K K
170 180 185
mCherry
cccgtgcagctgcccggcgcctacaacgtcaacatcaagttggacatcacctcccacaac
32
cccgtgcagctgcccggcgcctacaacgtcaacatcaagttggacatcacctcccacaac
P V Q L P G A Y N V N I K L D I T S H N
190 200 205
mCherry
gaggactacaccatcgtggaacagtacgaacgcgccgagggccgccactccaccggcggc
32
gaggactacaccatcgtggaacagtacgaacgcgccgagggccgccactccaccggcggc
E D Y T I V E Q Y E R A E G R H S T G G
210 220 225
mCherry
atggacgagctgtacaagggcggtaccgaattctaa
32
atggacgagctgtacaagggcggtaccgaattctaa
M D E L Y K G G T E F -
230


mCherry
47
ggtaccggcggctccatggtgagcaagggcgaggaggataacatggccatcatcaaggag
ggtaccggcggctccatggtgagcaagggcgaggaggataacatggccatcatcaaggag
G T G G S M V S K G E E D N M A I I K E
1 2 6 7 10
E
mCherry
47
ttcatgcgcttcaaggtgcacatggagggctccgtcaacggccacgagttcgagatcgag
ttcatgcgcttcaaggtgcacatggagggctccgtcaacggccacgagttcgagatcgag
F M R F K V H M E G S V N G H E F E I E
20 30
mCherry
47
ggcgagggcgagggccgcccctacgagggcacccagaccgccaagctgaaggtgaccaag
ggcgagggccagggccgcccctacgagggcacccagaccgccaagctgaaggtgaccaag
G E G Q G R P Y E G T Q T A K L K V T K
40 50
mCherry
47
ggtggcccctgcccttcgcctgggacatcctgtcccctcagttcatgtacggctccaag
ggtggcccctgcccttcgcctgggacatcctgtcccctcagttcatgtacggctccaag
G G P L P F A W D I L S P Q F M Y G S K
60 70
mCherry
47
gcctacgtgaagcaccccgccgacatccccgactacttgaagctgtccttccccgagggc
gcctacgtgaagcaccccgccgacatccccgactacttgaagctgtccttccccgagggc
A Y V K H P A D I P D Y L K L S F P E G
80 90
mCherry
47
ttcaagtgggagcgcgtgatgaacttcgaggacggcggcgtggtgaccgtgacccaggac
ttcaagtgggagcgcgtgatgaacttcgaggacggcggcgtggtgaccgtgacccaggac
F K W E R V M N F E D G G V V T V T Q D
100 110
mCherry
47
tcctccctgcaggacggcgagttcatctacaaggtgaagctgcgcggcaccaacttcccc
tcctccctgcaggacggcgagttcatctacaaggtgaagctgcgcggcaccaacttcccc
S S L Q D G E F I Y K V K L R G T N F P
120 130
mCherry
47
tccgacggccccgtaatgcagaagaagaccatgggctgggaggcctcctccgagcggatg
tccgacggccccgtaatgcagaagaagaccatgggctgggaggcctcctccgagcggatg
S D G P V M Q K K T M G W E A S S E R M
140 150
mCherry
47
taccccgaggacggcgccctgaagggcgagatcaagcagaggctgaagctgaaggacggc
taccccgaggacggcgccctgaagggcgagatcaagcagaggctgaagctgaaggacggc
Y P E D G A L K G E I K Q R L K L K D G
160 170
mCherry
47
ggccactacgacgctgaggtcaagaccacctacaaggccaagaagcccgtgcagctgccc
ggccactacgacgctgaggtcaagaccacctacaaggccaagaagcccgtgcagctgccc
G H Y D A E V K T T Y K A K K P V Q L P
180 190
mCherry
ggcgcctac----------aacgtcaacatcaagttggacatcacctcccacaac
G A Y N V N I K L D I T S H N
193 194 200 205
47
ggcgcctactgcggccgcacctacaacgtcaacatcaagttggacatcacctcccacaac
G A Y C G R T Y N V N I K L D I T S H N
mCherry
47
gaggactacaccatcgtggaacagtacgaacgcgccgagggccgccactccaccggcggc
gaggactacaccatcgtggaacagtacgaacgcgccgagggccgccactccaccggcggc
E D Y T I V E Q Y E R A E G R H S T G G
210 220 225
mCherry
47
atggacgagctgtacaagggcggtaccgaattctaa
atggacgagctgtacaagggcggtaccgaattctaa
M D E L Y K G G T E F -
230

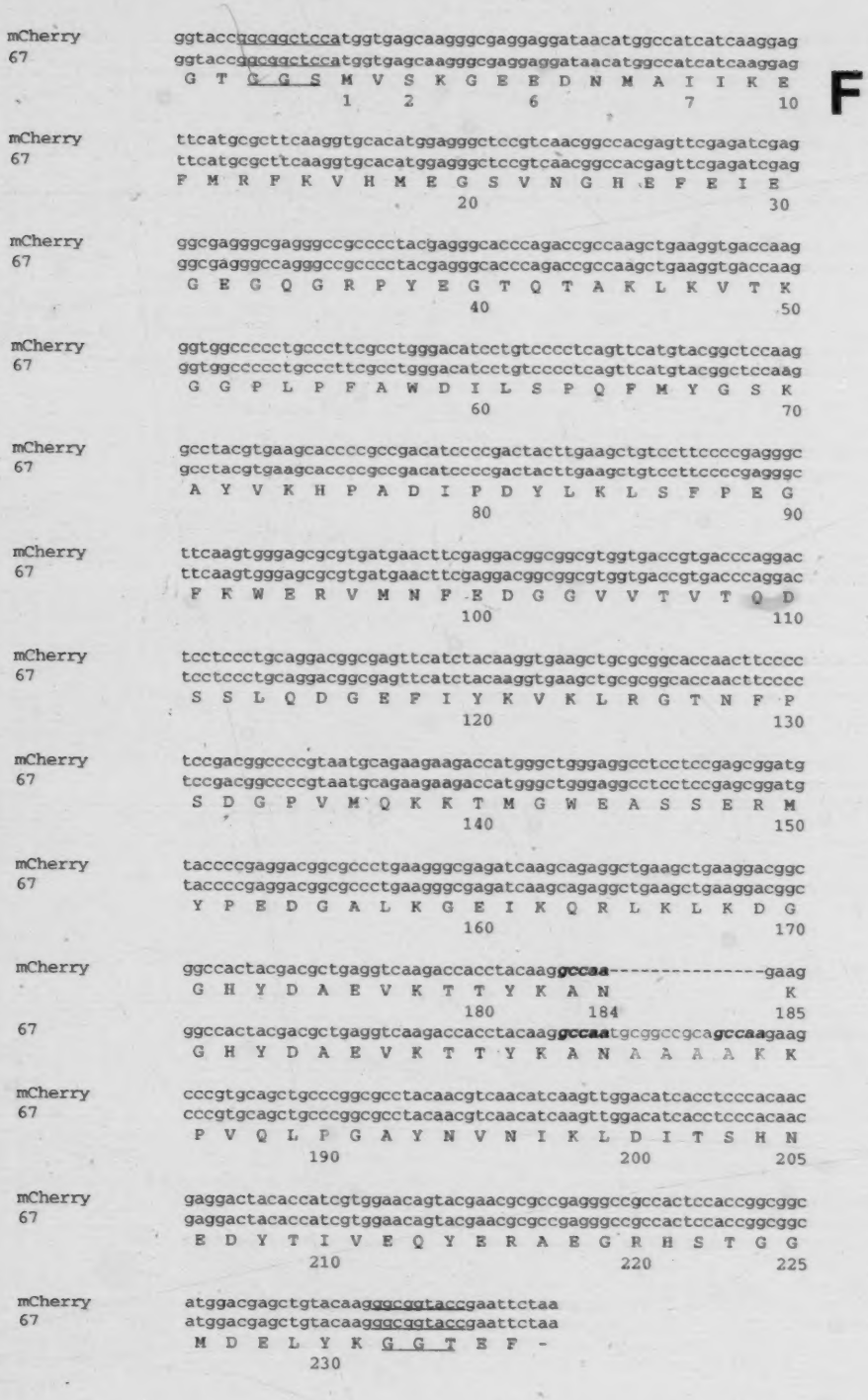

mCherry
67
G T G G S M V S K G E E D N M A I I K E
1 2 6 7 10
F
F M R F K V H M E G S V N G H E F E I E
20 30
G E G Q G R P Y E G T Q T A K L K V T K
40 50
G G P L P F A W D I L S P Q F M Y G S K
60 70
A Y V K H P A D I P D Y L K L S F P E G
80 90
F K W E R V M N F E D G G V V T V T Q D
100 110
S S L Q D G E F I Y K V K L R G T N F P
120 130
S D G P V M Q K K T M G W E A S S E R M
140 150
Y P E D G A L K G E I K Q R L K L K D G
160 170
G H Y D A E V K T T Y K A N K
180 184 185
G H Y D A E V K T T Y K A N A A A A K K
P V Q L P G A Y N V N I K L D I T S H N
190 200 205
E D Y T I V E Q Y E R A E G R H S T G G
210 220 225
M D E L Y K G G T E F -
230

**Figure 2.9** The sequence of the insertion mCherry variants compared to that of mCherry. 15 bp, translating to 5 amino acids, (gray) were inserted within mCherry, among which 5 bp (italics and bold) were duplicated from the gene of mCherry and the other 10 bp (tgcggccgca) remained from the Entranceposon, containing gcggccgc Not I restriction site. A, is #1 in which AAAQK was inserted between K138 and K139 within mCherry. B, is #2 in which MRPQQ was inserted between Q137 and K138 within mCherry. C, is #19 in which DAAAH was inserted between H25 and E26 within mCherry. D, is #32 in which AAASV was inserted between V22 and N23 within mCherry. E, is #47 in which CGRTY was inserted between Y193 and N194 within mCherry. F, is #67 in which NAAAA was inserted between A184 and K185 within mCherry.

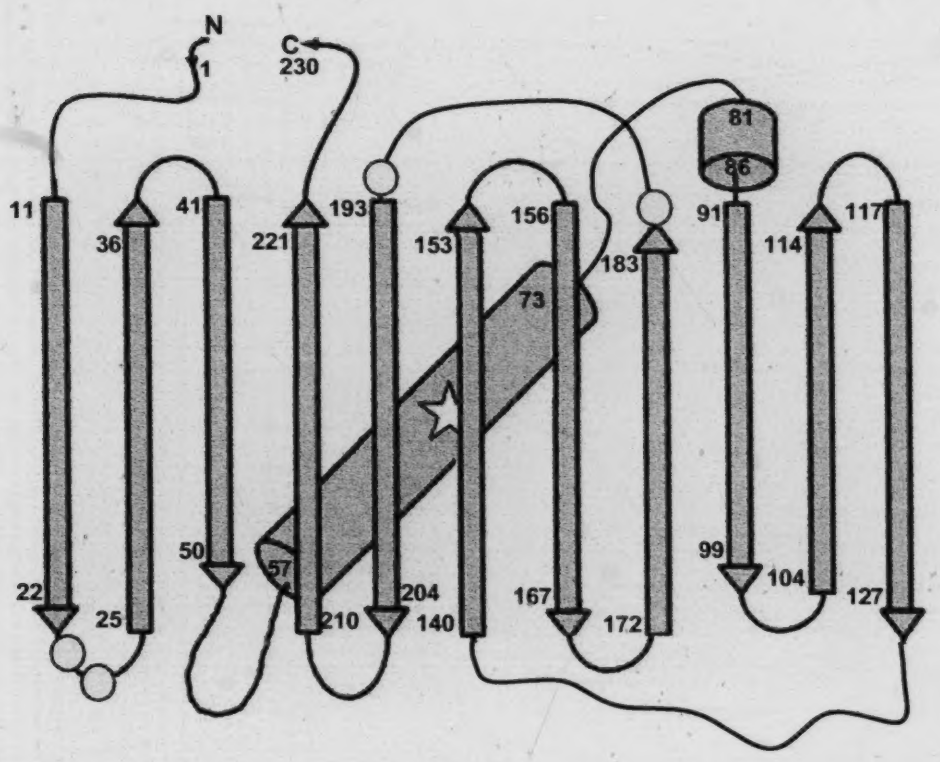



**Figure 2.10** Locations of 15 bp insertions in mCherry are shown in a schematic drawing of the overall fold of mCherry. Circles "O" are the insertion locations, the open star is the chromophore, residues 66-68, which is in the central helical segment (residues 58-72).

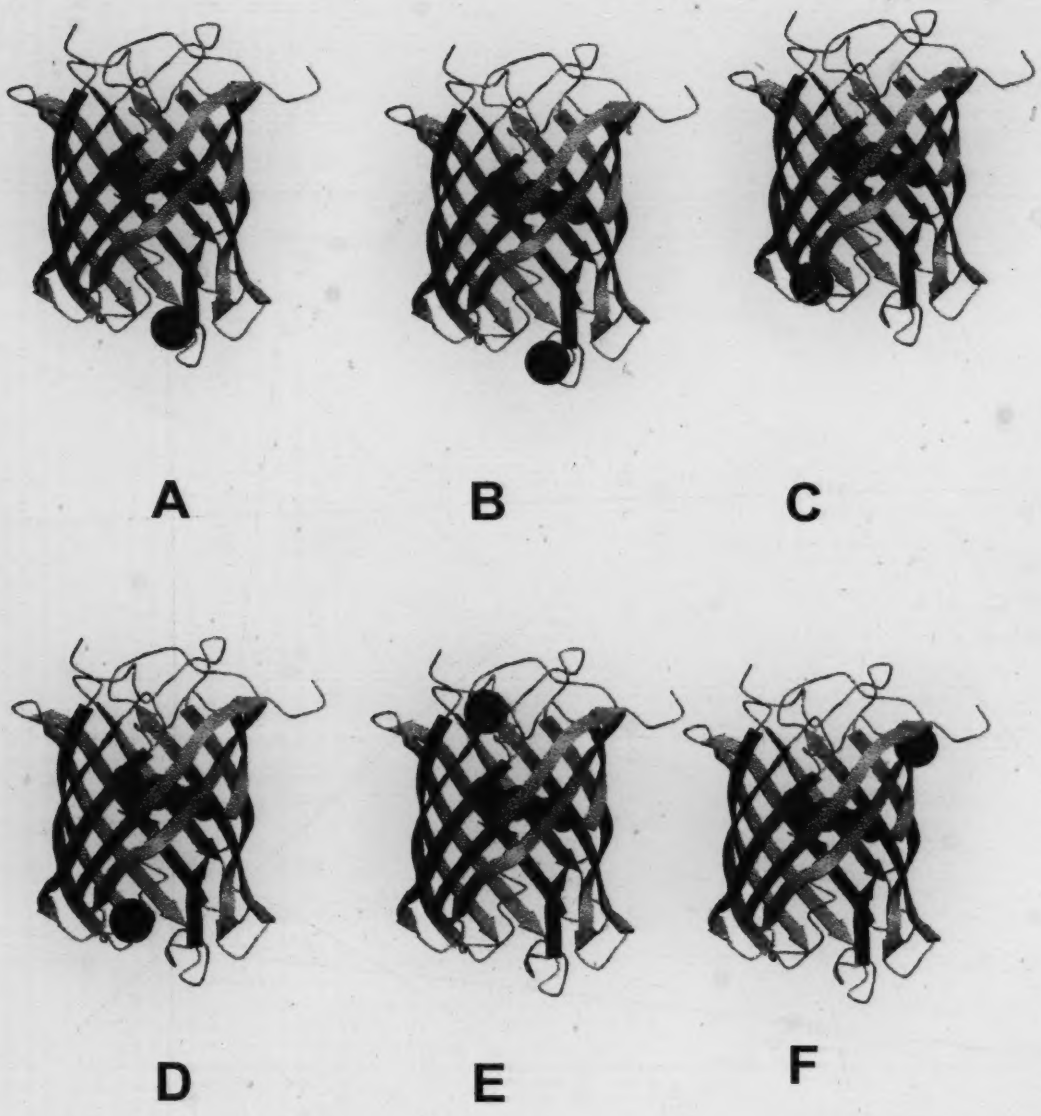



**Figure 2.11** Locations of 15 bp insertions in mCherry are shown in a cartoon drawing of mCherry. Balls are the insertion locations. A, B, C, D, E and F are #1 (138-139), #2 (137-138), #19 (25-26), #32 (22-23), #47 (193-194) and #67 (184-185) insertion mCherry variants.

### 2.3.3 CHARACTERIZATION OF INSERTION VARIANTS OF MCHERRY

The six mCherry variants derived from the 15 bp insertions exhibited the same spectra as mCherry [18]: an excitation maximum at 587 nm and an emission maximum at 610 nm. The fluorescence was measured with excitation at 550 nm (Figure 2.12). Compared with mCherry, all the insertion variants suffered some loss of fluorescent brightness to differing extents (Table 2.1), but kept the same peak shape. Among them, #67 (insertion between 184 and 185) showed least loss and was much brighter than the others. Only one peak (610 nm) showed in the spectra, so the insertion in these variants did not change the formation of the chromophore of the proteins. The decreasing of the fluorescence of the insertion variants might be the result of changing the environment of the chromophore of the proteins after the insertion of the five amino acids, such as quantum yield, extinction coefficient, protonation and folding efficiencies of the protein.

All six of the 15 bp insertion variants had fluorescence quantum yields (QY) (Table 2.1) and emission spectra (Figure 2.12) that were practically identical to that of mCherry. This result supports the conclusion that the immediate environment of the chromophore has not been significantly perturbed by the presence of the peptide insertions. One property in which the insertion variants are quite different from mCherry is in their 'in colony' fluorescence (Table 2.1). By imaging of colonies of bacteria on identically treated plates, we could assess the cumulative efficiency of all steps required to produce a functional FP; transcription, translation, folding, and chromophore formation. Given the nearly

identical genetic composition of the various insertion variants, there is unlikely to be significant differences in the efficiency with which the DNA is transcribed or the mRNA is translated. However, it is reasonable to suspect that peptide insertion could have dramatic effects on the ability of the protein to fold or the chromophore to develop. Indeed, each of our six insertion variants were decreased in 'in colony' fluorescent brightness (to varying extents) relative to mCherry suggesting that protein folding or chromophore maturation in each variant has been partially disrupted (Table 2.1).

The decreased 'in colony' brightness of the insertion variants relative to mCherry is partially explained by the diminished ε as provided Table 2.1. It is important to note that these apparent EC' values are an ensemble average and are based on the assumption that all protein molecules are equivalent. When we compared values of EC before and after alkali denaturation, we found that the ratio (intrinsic EC) of $EC_{native}$ to $EC_{denatured}$ for all variants was identical to that of mCherry. We interpret this result as evidence that the 'purified' proteins are actually mixtures of a red fluorescent protein (with brightness very similar to that of mCherry) and a non-fluorescent protein. This non-fluorescent protein may be a misfolded conformer that is incapable of promoting the series of post-translational modifications necessary for chromophore formation. Such a scenario is consistent with the observation that the quantum yields and emission peaks are also practically identical to mCherry. The corollary is that it is the relative concentration of these two components determines the ensemble EC'.

The decrease in apparent EC' due to the formation of a non-fluorescent species does not fully explain the decreased 'in colony' brightness. That is, the intrinsic brightness (QY * EC') determined for purified insertion variants are significantly greater percentages of the value for mCherry than the 'in colony' fluorescence (Table 2.1). We attribute this discrepancy to additional factors that are difficult to account for, such as the degree to which slow-folding variants may be targeted for degradation or inclusion body formation in *E. coli.*

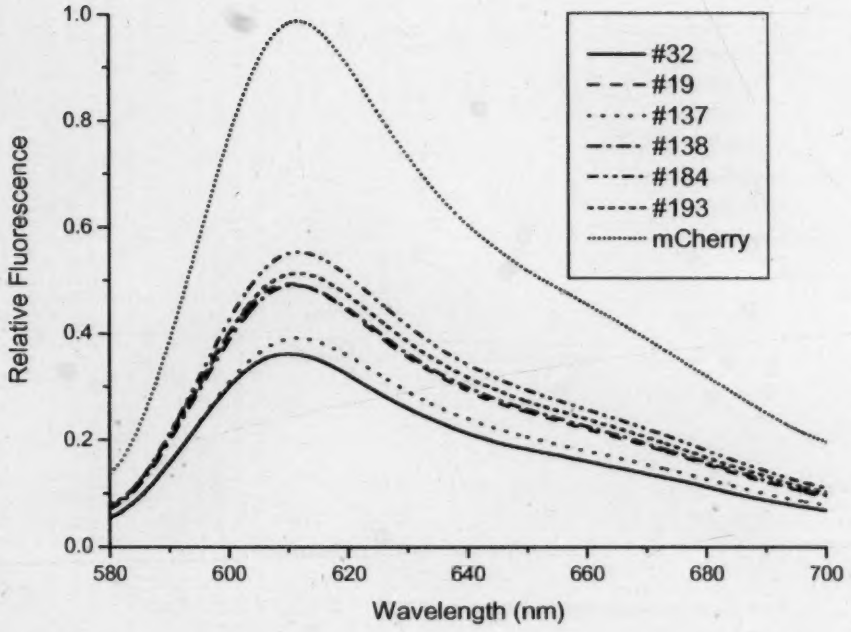



**Figure 2.12** The relative fluorescence spectral of purified insertion mCherry variants and mCherry. All the measurement were made under the same conditions and all the proteins had the same concentrations.

**Table 2.1** The characteristics of mCherry insertion variants compared to those of mCherry

| # | | #32 | #19 | #2 | #1 | #67 | #47 |
|---|---|---|---|---|---|---|---|
| Protein | mCherry | V22-AAASV-N23 | H25-DAAAH-E26 | Q137-MRPQQ-K138 | K138-AAAQK-K139 | K184-AAAQK-K185 | Y193-CGRTY-N194 |
| Relative Quantum Yield (QY) of Purified Protein | 100 % | 91% | 91% | 95 % | 82 % | 100 % | 100 % |
| Relative Intrinsic Extinction Coefficient (EC) * of Purified Protein | 100% | 100% | 110% | 110% | 90% | 100% | 100% |
| Relative Apparent Extinction Coefficient (EC') * of Purified Protein | 100 % | 36 % | 50 % | 40 % | 50 % | 55 % | 51 % |
| Relative Intrinsic Brightness (QY x EC') of Purified Protein | 100% | 33% | 46% | 38% | 41% | 55% | 51% |
| Relative Fluorescence of Bacterial colonies | 100% | 4% | 7% | 19% | 11% | 33% | 23% |

*: 280 nm was used to normalize the concentration to measure EC. EC was measured by base-denature.

## 2.4 CONCLUSIONS

The tightly interwoven, three-dimensional structure and intricate post-translation self-modification required for chromophore formation would prevent most insertions in FPs from maintaining fluorescence [14, 20]. Despite this limitation, in some certain locations five amino acids (encoded by a 15 bp gene) could be inserted within mCherry and the protein could still fold correctly, generate internal chromophores and protect the excited states from vibrational deactivation. Through an unbiased genetic screen, six sites (Figures 2.10) were found within the mCherry sequence that could tolerate insertion. All the locations are at a solvent-exposed loop between β-strands (Figure 2.11). These locations are 'permissive sites' that could potentially tolerate an entire protein insertion with an CR domain to make a Camgaroo sensor (Figure 1.1 B). The 'permissive sites' also are likely to tolerate the N- and C-termini of circular permuted variants. Among the six sites, the 183-184site might be the most promising to make the Camgaroo-type molecular sensors because it has the best ability to tolerate the insertion without serious disturbance of the mCherry structure and function. Although the 183-184 site is far from the chromophore (residues 66-68) in the primary sequence, it is three-dimensionally close to the chromophore, and thus is expected to be suitable for the purpose. The 137-138 or 138-139 sites might be promising for creation of circularly permuted variants, because they are at the opposite end of the β-barrel from the original N- and C- termini. The position of

137-138 or 138-139 within mCherry is similar to the position 144-145 within GFP, which is tolerant to the introduction of new termini of cpGFPs [14, 20].

## 2.5 LITERATURE CITED:


(1) Matz, M. V.; Fradkov, A. F.; Labas, Y. A.; Savitsky, A. P.; Zaraisky, A. G.; Markelov, M. L.; Lukyanov, S. A. *Nat. Biotechnol.* **1999**, *17*, 969-973.

(2) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509-544.

(3) Baird, G. S.; Zacharias, D. A.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 11984-11989.

(4) Campbell, R. E.; Tour, O.; Palmer, A. E.; Steinbach, P. A.; Baird, G. S.; Zacharias, D. A.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 7877-7882.

(5) Freimuth, P. I.; Taylor, J. W.; Kaiser, E. T. *J. Biol. Chem.* **1990**, *265*, 896-901.

(6) Sondek, J.; Shortle, D. *Proteins* **1992**, *13*, 132-140.

(7) Starzyk, R. M.; Burbaum, J. J.; Schimmel, P. *Biochemistry* **1989**, *28*, 8479-8484.

(8) Ladant, D.; Glaser, P.; Ullmann, A. *J. Biol. Chem.* **1992**, *267*, 2244-2250.

(9) Kavanaugh, J. S.; Moo-Penn, W. F.; Arnone, A. *Biochemistry* **1993**, *32*, 2509-2513.

(10) Heinz, D. W.; Baase, W. A.; Zhang, X. J.; Blaber, M.; Dahlquist, F. W.; Matthews, B. W. *J. Mol. Biol.* **1994**, *236*, 869-886.

(11) Doi, N.; Yanagawa, H. *Cell Mol. Life. Sci.* **1998**, *54*, 394-404.

(12) Doi, N.; Itaya, M.; Yomo, T.; Tokura, S.; Yanagawa, H. *FEBS Lett.* **1997**, *402*, 177-180.

(13) Doi, N.; Yanagawa, H. *FEBS Lett.* **1999**, *453*, 305-307.

(14) Baird, G. S.; Zacharias, D. A.; Tsien, R. Y. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11241-11246.

(15) Ai, H. W.; Henderson, J. N.; Remington, S. J.; Campbell, R. E. *Biochem. J.* **2006**.

(16) Cheng, Z.; Campbell, R. E. *Chembiochem* **2006**, *7*, 1147-1150.

(17) Brannon, J. H.; Magde, D. *J. Phys. Chem.* **1978**, *82*, 705-709.

(18) Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N.; Palmer, A. E.; Tsien, R. Y. *Nat. Biotechnol.* **2004**, *22*, 1567-1572.

(19) Chalfie, M.; Kain, S. *Green fluorescent protein: properties, applications, and protocols*; Wiley-Liss: New York, 1998.

(20) Topell, S.; Hennecke, J.; Glockshuber, R. *FEBS Lett.* **1999**, *457*, 283-289.

# CHAPTER 3

# CIRCULARLY PERMUTED VARIANTS OF MCHERRY AND PROGRESS TOWARDS GENETICALLY ENCODED CALCIUM SENSORS

## 3.1 INTRODUCTION

Circular permutation of FPs increases the flexibility and optical responsiveness to stresses by introducing new C- and N- termini [1]. The well-conserved β-can structure of FPs may appear to preclude the use of circular permutation. However, cpGFP, cpEGFP and cpYFP have been successfully created and applied. In Section 1.5.3 a thought experiment (Figure 1.20) was provided. In practice, this experiment is performed at the level of the genetic sequence encoding the corresponding protein chains. Normally, random (Figure 3.1 a, b, c and d) and rationally designed (Figure 3.1 a, b, e and f) circular permutations are made with the general method originally reported by Graf and Schachman [2].

The main use of these cpFPs is in fusing them to other proteins to render them fluorescent [3-7] at their flexible new termini. The properties of the circular permutation variants are somewhat altered with respect to the standard FPs and these new designs offer a unique opportunity to generate an original class of localization probes and physiological indicators[1].

Both the standard FPs and their cp counterparts have a rigid shell, composed of an 11-stranded β-barrel surrounding the chromophore. The shell not only protects them from photobleaching but also limits their environmental sensitivity. Therefore, many applications of mutagenesis or other forms of protein engineering have been applied to FPs and cpFPs to make biosensors that are

sensitive to their environments [8]. Normally, FPs or cpFPs have been fused to other proteins with a CR domain to make the molecular biosensors of interest.

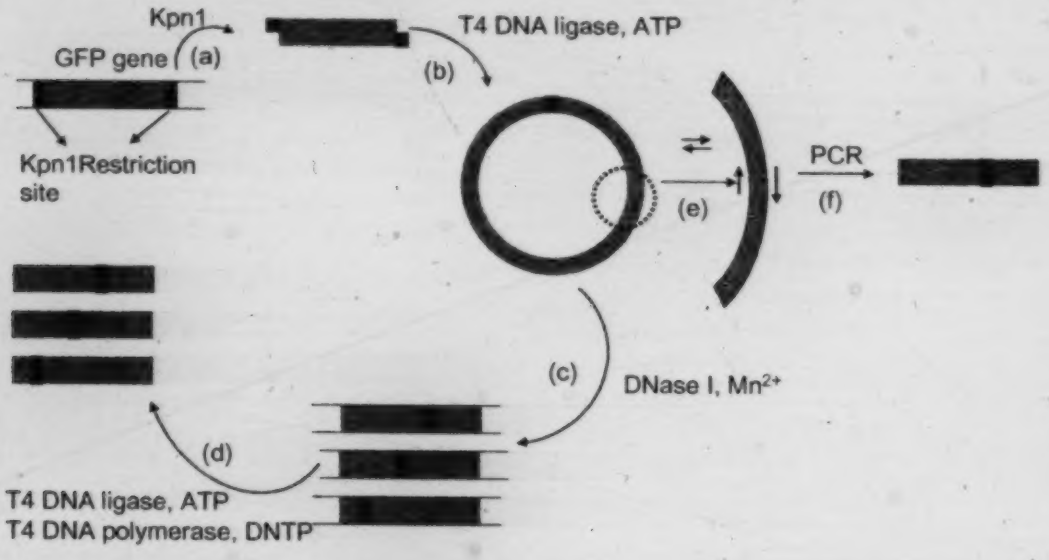



**Figure 3.1** Scheme for rational and random generation of circularly permuted variants of a protein. Both approaches utilize a circular gene that is constructed by cleavage at Kpn1 restriction sites flanking the natural gene (a) and subsequent ligation with T4 DNA-ligase (b). For random circular permutation, the circular gene is partially digested with DNase I (c) and a library of relinearized genes is isolated. The library is treated with T4 DNA ligase and T4 DNA polymerase (d), yielding blunt-ended, repaired fragments that can be cloned into an appropriate expression plasmid. The circular gene can also be used as a template for PCR amplification to generate rationally designed circularly permuted variants (e, f).

Typically $Ca^{2+}$ sensors based on FPs or cpFPs have been made and applied in many fields. Cytosolic and organellar free $Ca^{2+}$ concentrations are among the most important and dynamic intracellular signals. However, $Ca^{2+}$ fluctuations are often extremely localized and hard to measure. The signals have been most often measured using $Ca^{2+}$ sensitive fluorescent dyes[9-11], such as Fura-2 or Indo-1, or the bioluminescent protein, aequorin[12-14]. However, there are many disadvantages that are described in Section 1.3.2.

To overcome these problems, previous researchers developed the first dynamically responsive biochemical $Ca^{2+}$ indicators based on FPs, Cameleons (Figure 1.1A) [15, 16], which are proteins and thus can be genetically targeted to a specific location within the cell. These indicators are based on FRET between two green fluorescent proteins that are linked by a stretch of calmodulin (CaM); a protein that changes its shape in the presence of calcium and its binding peptide, M13.

Since they are genetically encoded, Cameleons have many advantages over previous small molecule based indicators. For example, they can be introduced into cells by DNA transfection rather than protein microconjection. In addition, they can readily be targeted to specific intracellular sites by fusing the corresponding cDNAs to appropriate organellar targeting signals or to localized host proteins to observe local $Ca^{2+}$ dynamics [17]. However, Cameleons, which are composed of two GFP-like molecules and a calcium binding protein, are fairly large and complex, which can in some cases significantly impair their targeting

efficiency [17]. Therefore, alternative calcium sensors have been developed that are based on a single GFP.

The fact that FPs can tolerate insertion and circular permutation [18, 19] led to the idea of creating new calcium biosensors to act as reporter proteins by inserting a calcium binding protein with a CR domain into FPs or cpFPs (Figures 1.1 B and C), or fusing the protein to the new termini of cpFPs (Figures 1.1 D). The first such example was the inserting of *Xenopus* calmodulin into position 145 of YFP [18], resulting in a calcium sensor named Camgaroo (Figure 1.1 B). Camgaroo increases fluorescence intensity sevenfold on binding $Ca^{2+}$ and has been used to measure calcium inside mitochondria and mushroom bodies in the brain of *Drosophila* [20]. Similarly, circularly permuted EGFP [3] and EYFP (V68L/Q69K) [5] have been fused between CaM and M13, the CaM–binding peptide of myosin light chain kinase, to create $Ca^{2+}$ sensors which have been termed Pericams (Figure 1.1 D) which are more sensitive than Camgaroo because the new termini of circular permutation increases the flexibility [1], the additional M13 incease the Ca2+ affinity, and the new termini are closer to the chromophore.

In both Camgaroos and Pericams, the binding of calcium to CaM that is fused to a single FP alters the fluorescent behavior of the FP's chromophore. So it was necessary to find a site in GFP where relatively long stretches of amino acids could be inserted without impairing its ability to fold into its characteristic cylindrical shape and form its chromophore.

In chapter 2, six locations within mCherry (Figure 2.10) were identified that permitted the insertion of a five-residue peptide (a fifteen bp in-frame

insertion) while maintaining red fluorescence. In this chapter, the use of these six sites to locate new N- and C- termini to generate new fluorescence-retaining circularly permuted variants of mCherry is described. In these variants the original N- and C- termini are joined by a flexible linker (Figure 1.21)

Circularly permuted variants of mCherry are a starting point for the creation of new genetically encoded indicators for use in live cell imaging. In addition, the favorable characteristics of mCherry, such as the long wavelength fluorescence, the good photostability, its fast maturation, its excellent pH resistance, and its high extinction coefficient [21], make mCherry a good probe choice to for generating genetically encoded indicators. These mCherry sensors will become the complementary biosensors of Camgaroo and Pericam GFP [3] of YFP [5, 6]..

## 3.2 EXPERIMENTAL

### 3.2.1 GENERAL METHODS AND MATERIALS

Please refer to Chapter 2.1.1.

### 3.2.2 GENE CONSTRUCTION

#### 3.2.2.1 Construction of Circularly Permuted mCherry

Circular permutations with a new N terminus at a specific site were made by two separate PCRs (Figures 3.1 a, b, e and f). The first PCR amplified mCherry gene constructed was Xba1-kpn1-GGS-mCherry-G-kpn1-EcoR1 gene (Figures 3.1 a), which has a 705 bp fragment encoding residues 1–230 of

mCherry and the hexapeptide linker and two *Kpn*I restriction sites. The gene was digested with XbaI and EcoRI and ligated into the similar plasmid vector, puc18 (Chapter 2.1.2). The mCherry plasmid DNA was digested with Kpn1 (Figure 3.1 b) and the resulting fragments were separated by standard agarose gel electrophoresis. The mCherry gene (about 700 bp in length) was extracted from the gel using QIAquick gel extract kit (Qiagen) according to the manufacturer's protocols and then self-ligated with T4 DNA ligase (Invitrogen). The ligation product, a circular gene (Figure 3.1 e) with a continuous reading frame was isolated and served as templates for the second PCR (Figure 3.1 f) for the amplification of the six circularly permuted mCherry genes with the oligonucleotide primers (Table 3.1). The PCR product was digested sequentially with Xba1 and EcoRI, and purified by QIAquick PCR clean-up kit (Qiagen) according to the manufacturer's protocols. The purified PCR fragment encoding the new N- and C- termini was ligated using T4 DNA ligase (Invitrogen) into pBAD/His B (Invitrogen) that had been digested with the same two restriction enzymes. Also, mCherry2 served as the template in the first PCR and made the cpmCherry2 as described above. All amplified genes were cloned into the pBAD/His B vector background and verified by dideoxynucleotide sequencing analysis.

#### 3.2.2.2 Construction of Camgaroo-type Gene

In Chapter 2, 15 bp with NotI restriction site were inserted within six different sites of the mCherry gene (Figure 2.10 and 2.11). The mCherry (with

**Table 3.1** The primers for the second PCR to construct the gene of circularly permuted mCherry (cpmCherry). The underlined letters show the duplicated gene at both the N- and C- temini of cpmCherry. The bold letters encode the stop codon gene and the italic letters are the restriction sites (In 5' primer is Xba1 and in 3' primer is EcoR1.) .

| sites | Residues corresponding to underlined codons | 5' primer | 3' primer |
|---|---|---|---|
| 21 | S21, V22 and N23 | CCGAG *TCTAGA* TCC GTG AAC GGCCACGAG | TTC *GAATTC* **TTA** GTT CAC GGA GCCCTCCATGTG |
| 25 | H25, E26 and F27 | CCGAG *TCTAGA* CAC GAG TTC GAGATCGAGGGC | TTC *GAATTC* **TTA** GAA CTC GTG GCCGTTCACGG |
| 136 | M136, Q137 and K138 | CCGAG *TCTAGA* ATG CAG AAG AAGACCATGGGC | TTC *GAATTC* **TTA** CTT CTG CAT TACGGGGCCGTC |
| 137 | Q137, K138 and K139 | CCGAG *TCTAGA* CAG AAG AAG ACCATGGGCTGG | TTC *GAATTC* **TTA** CTT CTT CTG CATTACGGGGCC |
| 183 | A183, K184 and K184 | CCGAG *TCTAGA* GCC AAG AAG CCCGTGCAG | TTC *GAATTC* **TTA** CTT CTT GGC CTTGTAGGTGGT |
| 192 | A192, Y193, N194 | CCGAG *TCTAGA* GCC TAC AAC GTCAACATCAAG | TTC *GAATTC* **TTA** GTT GTA GGC GCCGGGCAG |

insertions) plasmid DNA was digested with NotI (Invitrogen) and then treated with Bacterial Alkaline Phosphatase (BAP, Invitrogen) to remove 5′ phosphates. The resulting fragments were separated by standard agarose gel electrophoresis and extracted from the gel using QIAquick gel extract kit (Qiagen) according to the manufacturer's protocols. The extracted genes are generically referred to as mCherry vectors. Because of the random insertion (Chapter 2), the vectors have 3 different frames (Figures 3.2 A, C, D and E). The CaM/M13 gene was cloned from the gene encoding yellow cameleon3.3 (YC3.3) (a gift from Dr. Tsien's laboratory) by PCR amplification with primers that added a 5' NotI and a 3′ NotI site. Moreover, because the frames of the different mCherry vectors were different, additional bases were added after the NotI site in the 5′ primers to make the CaM/M13 gene in-frame and before the NotI site in 3′ primers to make the second part of mCherry gene not out of frame. No. 1, 19, 32 and 67 needed one additional base after the NotI site in the 5' primers and do not need any base before the NotI site in the 3' primers. No. 2 needed two additional bases after the NotI site in the 5' primer and needed two additional bases before the NotI site in the 3' primer. No. 47 did not need any additional base in either the 5' primer or the 3' primer. The PCR products were digested sequentially with NotI, and purified by QIAquick PCR clean up kit (Qiagen) according to the manufacturer's protocols. The purified PCR fragments were ligated using T4 DNA ligase (Invitrogen) into corresponding mCherry vectors.

One more PCR with Pfu polymerase (Fermentas) was performed to delete M13. The 5' end of the CaM DNA with a 5' primer coded for a Nar I site and the

01 = atg *cag* *aat* **gc^g gcc gca** *cag* *aag* aag acc atg = insert between K138 and K139

19 = ggc *cac* *gat* **gc^g gcc gca** *cac* *gag* ttc gag atc = insert between H25 and E26

32 = gag ggc *tcc* *gtt* **gc^g gcc gca** *tcc* *gtg* aac ggc = insert between V22 and N23

67 = aag gcc *aat* **gc^g gcc gca** *gcc* *aag* aag ccc gtg = insert between A183 and K184

02 = atg *cag* **atg c^gg ccg cag** *cag* *aag* aag acc atg = insert between Q137 and K138

47 = ccc ggc *gcc* *tac* **tgc^ ggc cgc acc** *tac* aac gtc = insert between Y193 and N194

**A**

ggc cgc **CaM/M13** gc

**B**

gc g gcc

**C**

c gg cc

**D**

tgc ggc c

**E**

**Figure 3.2** (A) shows the sequence insertion in mCherry gene. 10 bp (bold) remnant of the Entranceposon, containing gcggccgc NotI restriction site. 5 bp (italics) were duplicated from the gene of mCherry. ^ is the digestion site. (B) is the gene of CaM/M13 after digestion with NotI. (C) is the mCherry vector of No. 1, 19, 32 and 67 after digestion with NotI without the additional bases. (D) is the mCherry vector of No. 2 after digestion with NotI. (E) is the mCherry vector of No. 47 after digestion with NotI.

3' start of the second part of mCherry coded for a Nar I site. The PCR fragment was purified by QIAquick PCR clean up kit (Qiagen). The purified PCR fragment was self-ligated with T4 DNA ligase (Invitrogen).

### 3.2.2.3 Construction of Pericam Gene

Two PCRs were performed to prepare the CaM/M13 vector. The first PCR amplified CaM/M13 gene and constructed the XbaI-CaM/M13-EcoRI gene from YC3.3 and was purified by QIAquick PCR clean up kit (Qiagen). The purified PCR fragment was digested with XbaI and EcoRI and ligated into the pBAD/His B vector digested with the same restriction enzyme. Electrocompetent *Escherichia coli* strain DH10B (Invitrogen) was transformed, plated on LB/agar plates and pooled into medium as described previously. The plasmid DNA was prepared by the QIAquick miniprep kit (Qiagen) according to the manufacturer's protocols as the template of the second PCR. The 5' start of the M13 with a 5' primer coded for start codon and SphI site and the 3' end of the CaM with 3' primer coded for stop codon and XhoI site. The PCR product was digested sequentially with SphI and XhoI, and purified by QIAquick PCR clean up kit (Qiagen). The purified PCR fragment was ligated using T4 DNA ligase (Invitrogen) with the digested cpmCherry with same restriction sites, which were prepared by another PCR with the 5' primer coding XhoI and the 3' primer coding SphI.

### 3.2.3 BACTERIAL TRANSFORMATION AND PROTEIN PURIFICATION

*Escherichia coli* LMG194 Gold (Invitrogen) was transformed by electroporation in 10% glycerol with a ligation mixture (0.2 cm cuvette, 12.5 kV/cm, 200Ω, 25 μF) and was plated on LB/agar containing ampicillin (0.1 mg/mL) and arabinose (0.2%). A single colony was used to inoculate a 4 mL culture that was allowed to grow overnight (37 °C, 225 rpm) before being diluted into 1 L of LB medium containing ampicillin (0.1 mg/mL) and arabinose (0.2%). For protein expression, the culture was allowed to grow at 37°C for about 4 h to let cells grow to an $OD_{600}$ of 0.6 and then grow at 18 °C for 48 h before cells were harvested by centrifugation and lysed by French Press. In parallel, the same culture kept at 37 °C overnight to compare the maturation of the chromophore of the cpmCherry. The bacteria were then pelleted by centrifugation, resuspended in 50 mM Tris–HCl/300 mM NaCl, and lysed by a French press. The bacterial lysates were centrifuged at 30,000 × *g* for 30 min, and the proteins were purified from the supernatants by Ni-NTA chromatography (Amersham).

### 3.2.4 LIBRARY CONSTRUCTION, MUTAGENESIS AND LIBRARY SCREENING

Randomly mutated libraries were constructed by error-prone PCR as previously described [22, 23] using Taq polymerase (New England Biolabs) under conditions optimal for 3 mutations per 1,000 bp [24]. Full-length gene libraries resulting from error-prone PCR were digested with Xba1 and EcoR1 and ligated into similarly digested pBAD/His B vector. Electrocompetent *Escherichia coli* strain DH10B (Invitrogen) was transformed and plated on LB/agar plates as described above. Plates were incubated for 14 h at 37 °C prior to screening.

The bacterial colony screening is given in chapter 2.2.4.

### 3.2.5 SPECTROSCOPY OF THE CIRCULARLY PERMUTED MCHERRY VARIANTS

The recording of the absorption and emission spectra is given in chapter 2.2.7.

## 3.3 RESULTS AND DISCUSSION

### 3.3.1 CONSTRUCTION OF CIRCULARLY PERMUTED VARIANTS OF MCHERRY

The original N- and C- termini of mCherry were connected with the hexapeptide linker GGTGGS. The fluorescent protein sequences were begun at a specific site x, (example: 183A) as the new N- terminus, and x+2 (example: 185K) became the new C- terminus of the circularly permuted variants of mCherry (Figure 3.3). Three amino acids were duplicated at both new N- and C- termini of cpmCherry variants to help the new termini fold properly. Figure 3.3 shows the positions of the new termini in the six circularly permuted variants that were constructed and investigated in this study. The six sites were chosen based on the insertions within mCherry that maintained fluorescence that are documented in Chapter 2. In all of these circular variant constructs, new termini were introduced into the loops at the top and bottom of the β-barrel of mCherry.

From examination of this model (Figure 3.3), only the 22-23, 25-26, 137-138 or 138-139 positions might be the the most potentially interesting of the circularly permuted proteins. The reason is that the new N- and C- termini are at the opposite end of the β-barrel from the wild-type protein [18, 19]. However, none of these circular permutation mutants at these sites stayed fluorescent, even after one round of random mutagenesis. The circular permutations at 183-184 or 192-193 have new N- and C- termini at the same end of the β-barrel as the wild type protein and do exhibit significant fluorescence.

### 3.3.2 SPECTROSCOPIC PROPERTIES OF THE CIRCULARLY PERMUTED VARIANTS

For further characterization of the fluorescent circularly permuted variants, the spectroscopic properties of the purified fluorescent variants were compared with mCherry at identical protein concentrations. The cpmCherry variants at the 21, 25, 136 and 137 sites were non-fluorescent, even when expressed at 18 °C, but the cpmCherry variants at 183 (183cpmCherry) and 192 (192cpmCherry) sites did exhibit fluorescence. When the excitation was 480 nm and 540 nm the spectra (Figure 3.4 A) of both 183cpmCherry and 192cpmCherry exhibit broad emission peaks with maxima at 516 nm (green peak) and 610 nm (red peak), respectively. In contrast, mCherry [21] (Figure 3.4 B) and the insertion mCherry in Chapter 2, exhibit only one very sharp emission spectrum maximum at 610 nm for all excitation wavelengths. The wild type (WT) DsRed [25] exhibits a very broad emission spectrum (spectral width about 50-60 nm) that peaks at 570-610 nm and has a maximum at 579 nm.

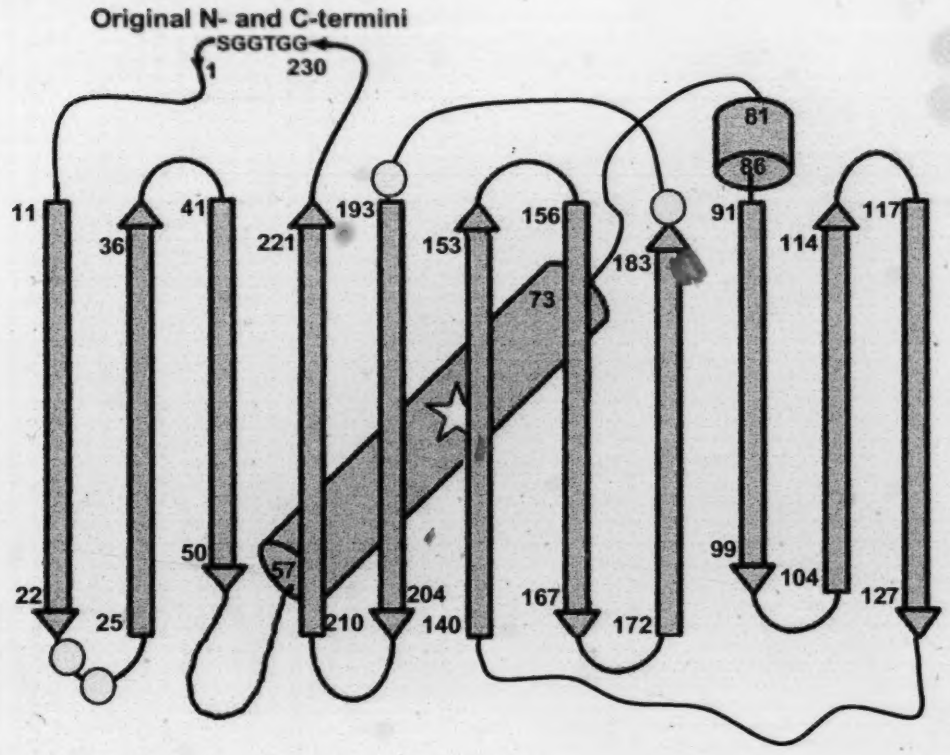



**Figure 3.3** Locations of new termini in fluorescent, circularly permuted mCherrry variants are shown in: a schematic drawing of the overall fold of mCherry modified to show the starting points of fluorescent circular permutations (O), the linker (GGTGS) connecting the original N- and C- termini, and the approximate location of the chromophore (open star residues 66-68), which is in the central helical segment (residues 58-72).

In order to improve the cpmCherry, random mutagenesis was performed. After one round of random mutagenesis by error-prone PCR, the non-fluorescent variants of cpmCherry permuted at positions 21, 25, 136 or 137 were identified. However, after random mutagenesis, brighter variants of 192cpmCherry (Figure 3.5) and 183cpmCherry (Figure 3.6) were identified. These variants are referred to as 192cpmCherry1.2 (equivalent to 192cpmCherry A57V, E39Q) and 183cpmCherry1.5 (equivalent to 183cpmCherry N195D, N6bI). In these variants, the intensity of the green peak was suppressed and the intensity of the red peak was enhanced.

In addition, the spectra of the cpmCherry expressed at 18 °C had a higher intensity of red emission peak but lower intensity of green peak than the protein expressed at 37 °C (Figures 3.5 and 3.6). That suggests that the circular permuted mCherry variants can mature more completely but slower at lower temperature. The behavior of the cpmCherry is similar to wild-type GFP [26] and circularly permuted GFP [18] which mature more completely at low temperature (4 °C in the literature [19]), and unlike the DsRed which matures faster and completely at 37 °C [27].

After the *in vitro* cpmCherry was stored at 4 °C for months to let the protein completely maturate, the spectra were measured. The spectra (Figures 3.7 and 3.8) shows that the intensity of the red peak was enhanced and the green peak was suppressed greatly. The cpmCherry can continue maturation *in vitro* slowly at low temperature (4 °C), making the protein fold better and the maturation more complete.

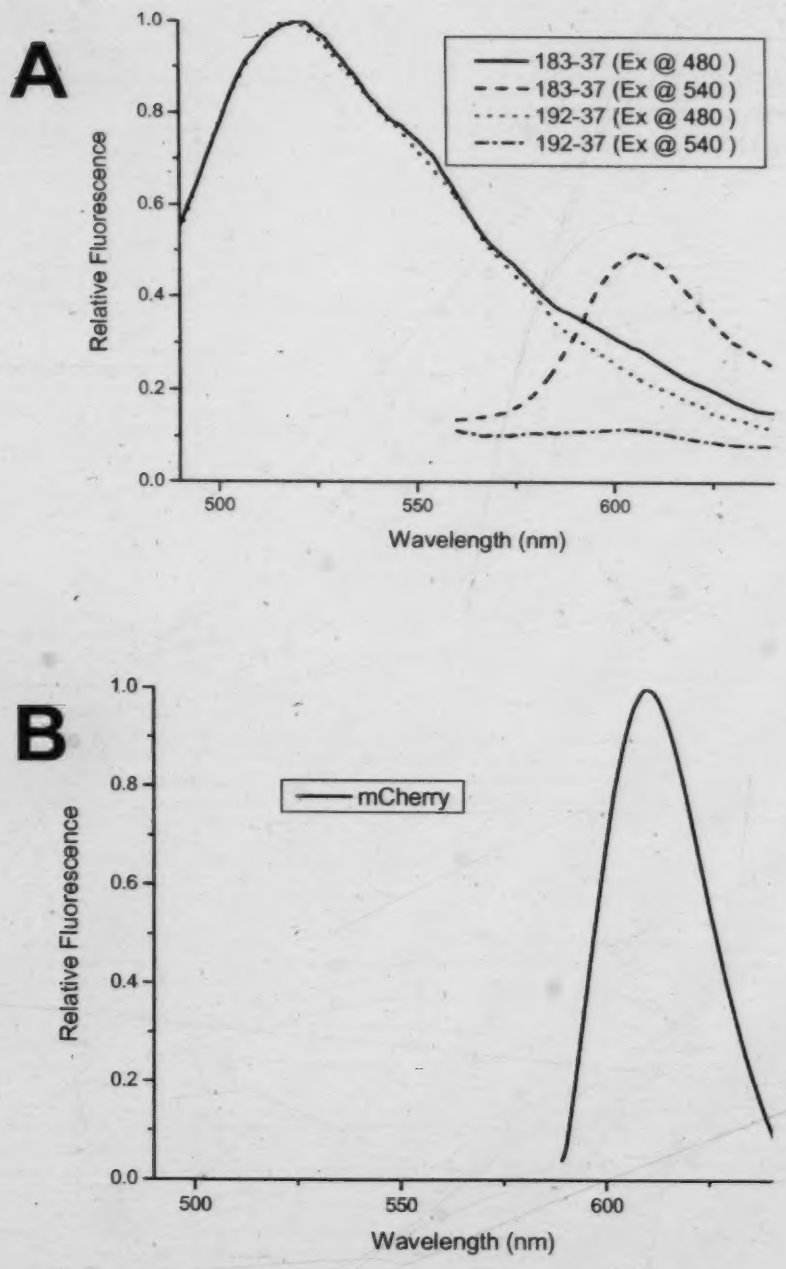



**Figure 3.4** (A)The emission spectra of 183cpmCherry and 192cpmCherry, which were expressed at 37 °C. The excitation was 480 nm (labeled Ex @ 480) and 540 (labeled Ex @ 480) separately. The spectra were normalized based on the green peak of the spectra excited at 480nm. (B) The emission spectra of mCherry.

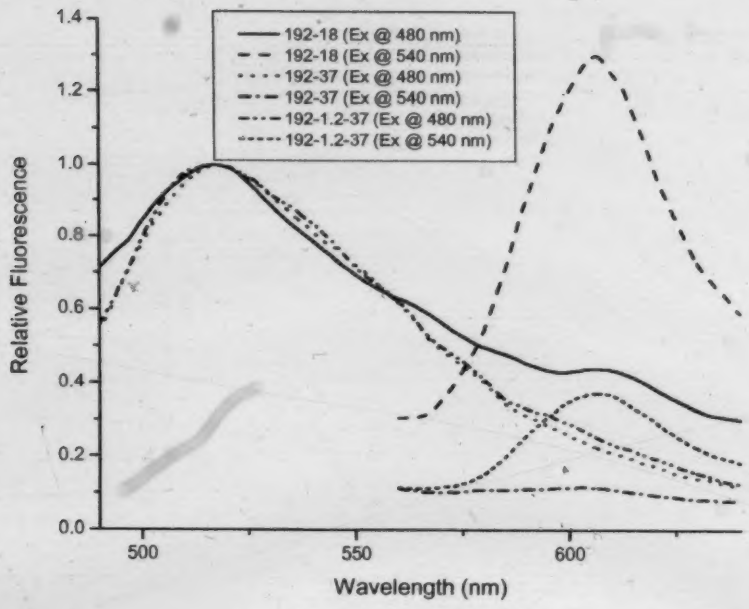



**Figure 3.5** The emission spectra of 192cpmCherry. 192-18 is 192cpmCherry expressed at 18 °C; 192-37 is 192cpmCherry expressed at 37 °C; and 192-1.2-37 is 192cpmCherry1.2 expressed at 37 °C. The spectra were normalized based on the green peak of the spectra excited at 480nm.

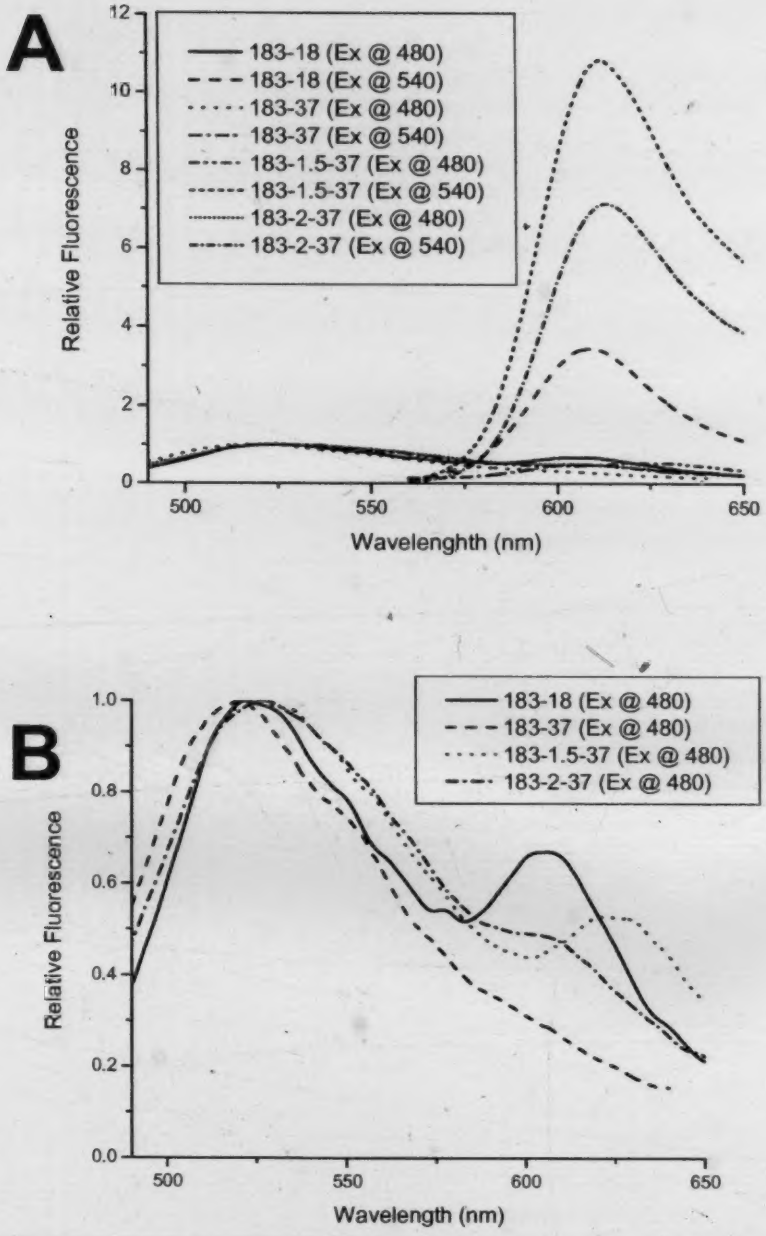



**Figure 3.6** The emission spectra of 183cpmCherry and 192cpmCherry. 183-18 is 183cpmCherry expressed at 18 °C; 183-37 is 183cpmCherry expressed at 37 °C; 183-M-37 is 183cpmCherry1.5 expressed at 37 °C; and 183-1.5-37 is the mutated 183 circularly permuted mCherry2 at 183 site expressed at 37 °C. The spectra were normalized based on the green peak of the spectra excited at 480 nm.

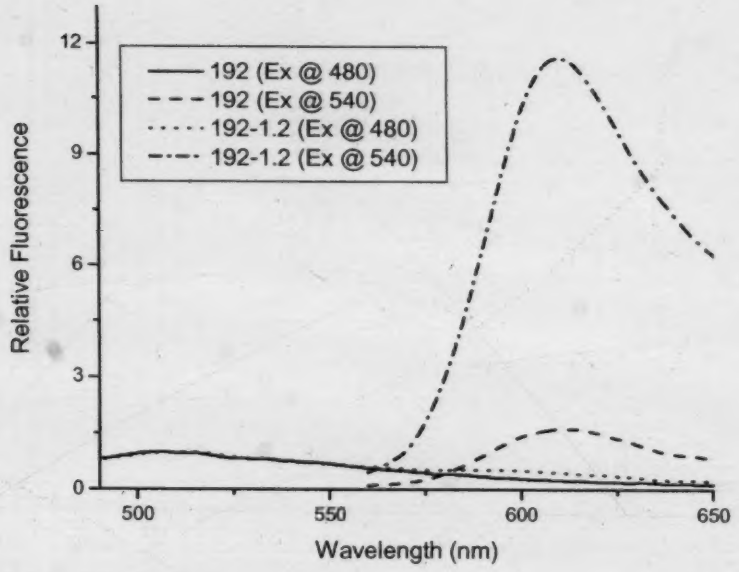
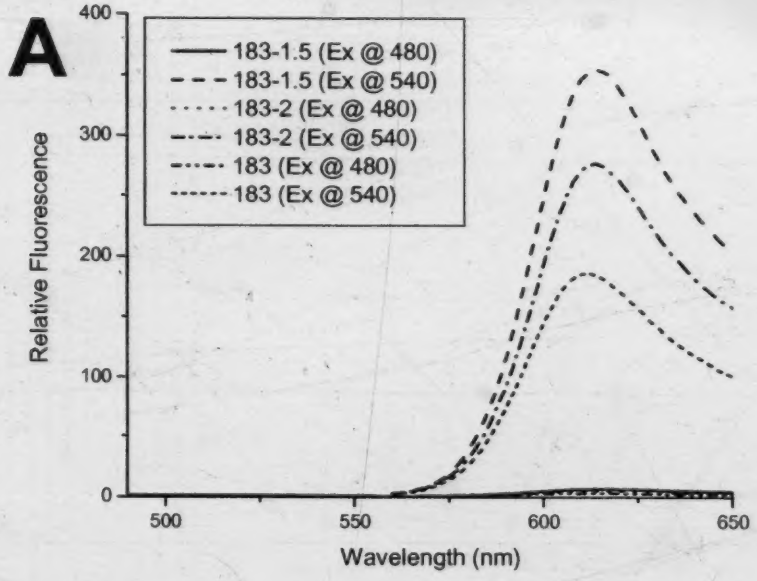



**Figure 3.7** The emission spectra of 192cpmCherry variants expressed at 37 °C. 192-1.2-37 is 192cpmCherry1.2. The spectra were normalized based on the green peak of the spectra excited at 480nm.



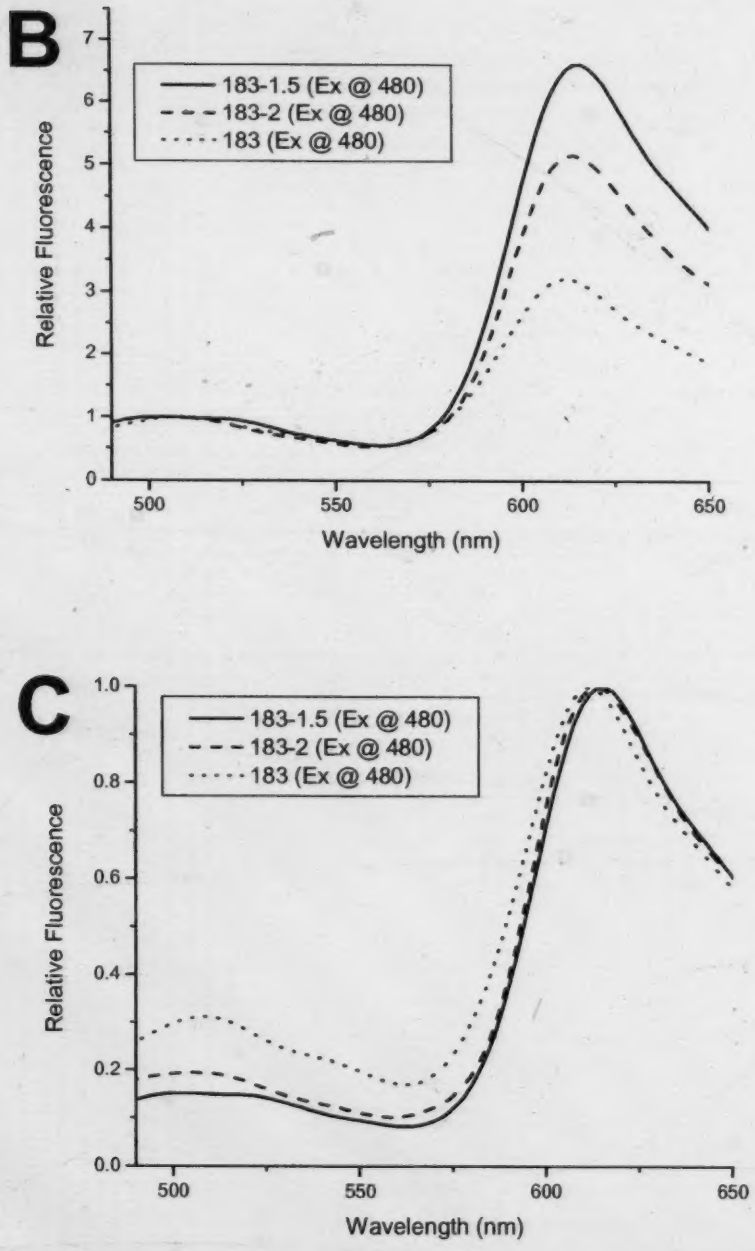



**Figure 3.8** The *in vitro* emission spectra of 183cpmCherry variants expressed at 37 °C and stored at 4 °C for months. 183-1.5 is 183cpmCherry1.5 and 183-2 is the circularly permuted mCherry2 at 183 site. (A) and (B) were normalized based on the green peak of the spectra excited at 480nm. (C) was normalized based on the red peak of the spectra excited at 480nm.

For further characterization of the Camgaroos, the spectra of mCherry with CaM/M13 insertions were measured. Unfortunately, the proteins are not fluorescent in the presence of either EDTA or $Ca^{2+}$. The construction of mCherry with a CaM only insertion (Camgaroo) or the Pericam-type fusion is still in process.

## 3.4 CONCLUSIONS

Circular permutation is known [28] to be tolerated in a variety of proteins when the original N- and C- termini are in close proximity, as is true for FPs, including mCherry. It is easy to imagine that the function of a protein could be preserved despite the rearrangement of the sequence of the protein, if the protein consists of more than one autonomous domain loosely held together by flexible linkers. At first glance, FPs would seem poor candidates because of their monolithic cylindrical symmetry, intricate interdigitated β-barrels made up by 11 β-sheets, complexity of their maturation process, and intolerance of significant N- and C-terminal truncations [29]. Nevertheless, here two different sites within the mCherry sequence at which a circularly permuted fluorescent protein could retain fluorescence were found. However, the circularly permuted mCherry variants investigated in this study show that mCherry exhibits a lower tolerance towards circular permutations than other fluorescent proteins [3-6, 19, 30, 31].

Baird et. al. [18] have summarized the many topologies for merging the sequences of a host protein and a FP or cpFP (Figure 3.9). A183 and A192 are plausible sites at which proteins may be inserted (as in a Camgaroo, Figure 3.9 D)

and also may be preferred sites to insert the cpmCherry sequence into the host (as in a Pericam, Figure 3.9 G). The sensitivity of cpmCherry variants to the environment and to small changes in sequence around new termini ought to enhance the probability that their fluorescence will sense changes in the conformations of proteins into which they have been spliced. Therefore, it is predicted that the insertion within cpmCherry (Figure 3.9 G) should confer greater responsiveness than the insertion in mCherry (Figure 3.8 D). The cpmCherry should provide a very useful choice for multicolor imaging of fusion proteins.

Although the mCherry with CaM/M13 insertion (like Figure 3.9 D, Camgaroo) is not fluorescent, even with $Ca^{2+}$ added, mutagenesis can be used to improve mCherry with CaM only insertion. Also Pericams might be more promising because the new termini ought to enhance the probability that the fluorescence will sense changes in conformations of proteins into which they have been spliced.

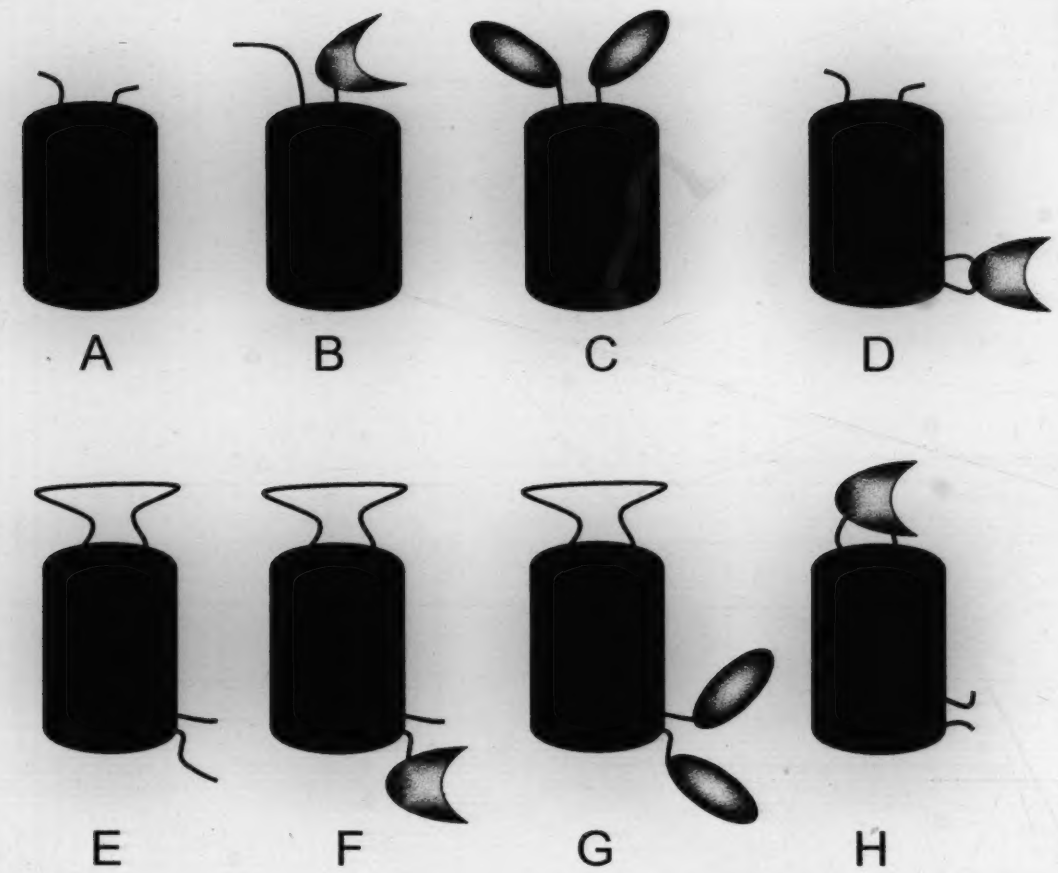



**Figure 3.9** Topologies of FP, cpFP, and chimeras with other proteins. The other proteins are depicted schematically as moon and ellipse. (A) The original FP molecule is represented as a cartoon drawing; (B) A tandem fusion product of FP with another protein; (C) the insertion of FP into a split protein domain; (D) insertion of a complete protein domain into the β-barrel backbone of FP; (E) cpFP connected the original termini with a loop and produced new termini; (F) A tandem fusion product of cpFP with another protein; (G) the insertion of cpFP into a split protein domain; (H) insertion of a complete protein domain into the β-barrel backbone of cpFP.

## 3.5 LITERATURE CITED


(1) Zhang, J.; Campbell, R. E.; Ting, A. Y.; Tsien, R. Y. *Nat. Rev. Mol. Cell. Biol.* **2002**, *3*, 906-918.

(2) Graf, R.; Schachman, H. K. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 11591-11596.

(3) Nagai, T.; Sawano, A.; Park, E. S.; Miyawaki, A. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 3197-3202.

(4) Shimozono, S.; Fukano, T.; Nagai, T.; Kirino, Y.; Mizuno, H.; Miyawaki, A. *Sci STKE* **2002**, *2002*, PL*A*.

(5) Nagai, T.; Yamada, S.; Tominaga, T.; Ichikawa, M.; Miyawaki, A. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 10554-10559.

(6) Belousov, V. V.; Fradkov, A. F.; Lukyanov, K. A.; Staroverov, D. B.; Shakhbazov, K. S.; Terskikh, A. V.; Lukyanov, S. *Nat. Methods* **2006**, *3*, 281-286.

(7) Chiang, J. J.; Li, I.; Truong, K. *Biotechnol. Lett.* **2006**, *28*, 471-475.

(8) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509-544.

(9) Grynkiewicz, G.; Poenie, M.; Tsien, R. Y. *J. Biol. Chem.* **1985**, *260*, 3440-3450.

(10) Hofer, A. M.; Schulz, I. *Cell Calcium* **1996**, *20*, 235-242.

(11) Golovina, V. A.; Blaustein, M. P. *Science* **1997**, *275*, 1643-1648.

(12) Rutter, G. A.; Burnett, P.; Rizzuto, R.; Brini, M.; Murgia, M.; Pozzan, T.; Tavare, J. M.; Denton, R. M. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 5489-5494.

(13) Knight, M. R.; Read, N. D.; Campbell, A. K.; Trewavas, A. J. *J. Cell Biol.* **1993**, *121*, 83-90.

(14) Miller, A. L., Karplus, E., and Jaffe, L.F. *Practical Guide to the Study of Calcium in Living Cell*: New York, 1994.

(15) Miyawaki, A.; Llopis, J.; Heim, R.; McCaffery, J. M.; Adams, J. A.; Ikura, M.; Tsien, R. Y. *Nature* **1997**, *388*, 882-887.

(16) Romoser, V. A.; Hinkle, P. M.; Persechini, A. *J. Biol. Chem.* **1997**, *272*, 13270-13274.

(17) Hicks, B. W.; Humana Press, 2002.

(18) Baird, G. S.; Zacharias, D. A.; Tsien, R. Y. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11241-11246.

(19) Topell, S.; Hennecke, J.; Glockshuber, R. *FEBS Lett.* **1999**, *457*, 283-289.

(20) Yu, D.; Baird, G. S.; Tsien, R. Y.; Davis, R. L. *J. Neurosci.* **2003**, *23*, 64-72.

(21) Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N.; Palmer, A. E.; Tsien, R. Y. *Nat. Biotechnol.* **2004**, *22*, 1567-1572.

(22) Campbell, R. E.; Tour, O.; Palmer, A. E.; Steinbach, P. A.; Baird, G. S.; Zacharias, D. A.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 7877-7882.

(23) Ai, H. W.; Henderson, J. N.; Remington, S. J.; Campbell, R. E. *Biochem. J.* **2006**.

(24) Fromant, M.; Blanquet, S.; Plateau, P. *Anal. Biochem.* **1995**, *224*, 347-353.

(25) Matz, M. V.; Fradkov, A. F.; Labas, Y. A.; Savitsky, A. P.; Zaraisky, A. G.; Markelov, M. L.; Lukyanov, S. A. *Nat. Biotechnol.* **1999**, *17*, 969-973.

(26) Chalfie, M. *Photochem Photobiol* **1995**, *62*, 651-656.

(27) Yarbrough, D.; Wachter, R.; Kallio, K.; Matz, M.; Remington, S. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 462-467.

(28) Heinemann, U.; Hahn, M. *Prog. Biophys. Mol. Biol.* **1995**, *64*, 121-143.

(29) Dopf, J.; Horiagon, T. M. *Gene* **1996**, *173*, 39-44.

(30) Akemann, W.; Raj, C. D.; Knopfel, T. *Photochem. Photobiol.* **2001**, *74*, 356-363.

(31) Wang, T.; Sakai, Y.; Nakajima, K.; Miyawaki, A.; Ito, K.; Hara, M. *Colloids Surf B Biointerfaces* **2005**, *40*, 183-187.

# CHAPTER 4:

# CONCLUSIONS AND FUTURE WORK

Recently, both green and yellow fluorescent proteins (GFPs and YFPs) have been genetically engineered to create circular permutations and have been further developed to create genetically encoded $Ca^{2+}$ sensors. In this work we have taken the first steps towards the creation of an analogous red $Ca^{2+}$ sensor based on a monomeric RFP known as mCherry.

The monolithic FPs, which are tightly interwoven by an 11-stranded β-barrel, give us a picture of a protein that appears to have no ability to tolerate circular permutations and insertions. However, in this work we have demonstrated that mCherry can tolerate circular permutations and insertions. In addition, the perfectly rigid shell in FPs surrounding the chromophore generally hinders environmental sensitivity. This thesis describes the work that is in progress towards making $Ca^{2+}$ sensors by protein engineering.

In Chapter 2, five amino acids (encoded by a 15 bp gene fragment) were inserted in six specific sites within mCherry producing mutants that could still fold very well, generate internal chromophores and retain their red fluorescence. All of these insertion sites are located at a solvent-exposed loop between β-strands of mCherry, which are likely to be "permissive sites" that tolerate the new N- and new C-termini of a circular permuted variants or tolerate an entire protein insertion to make Camgaroo sensors. Among these sites, 183-184 might be the most promising to tolerate a protein insertion to make a Camgaroo or make the circularly permuted mCherry and further to make a Pericam because it has the best ability to tolerate an insertion and maintain bright fluorescence. Although 184-184 is far from the chromophore (residues 66-68) in the primary sequence, it

is three-dimensionally close to the chromophore and thus is expected to be suitable for the purpose.

Like other protein candidates tolerant of circular permutation, mCherry has the original N- and C- termini in close proximity. This feature is generally accepted to be a prerequisite for successful circular permutation. In Chapter 3, two fluorescent circularly permuted mCherry (cpmCherry) mutants were constructed at different sites (183 and 192) within the mCherry sequence. Moreover, these cpmCherry variants were improved by means of random mutagenesis by error-prone PCR. The cpmCherry variants have emission peaks at both 516 nm and 610 nm. In this regard they differ from the original mCherry that has only one emission peak at 610 nm, and the WT DsRed with an emission peak at at 579 nm. However, after random mutagenesis by error-prone PCR, the spectra of the cpmCherry variants at both 183 (N195D, N6bI) and 192 (A57V, E39Q) suppressed the intensity of the green peak and enhanced the intensity of the red peak. At the same time, the cpmCherry could maturate more quickly and completely at a lower temperature: 18 °C instead of 37 °C.

The cpmCherry is a good basis from which to make $Ca^{2+}$ biosensors: Camgaroo-like (Figure 1.1 B and Figure 3.8 D) or Pericam-like (Figure 1.1 D and Figure 3.8 G) biosensors. Although green fluorescent single FP $Ca^{2+}$ biosensors are known, a comparable red biosensor is conspicuously absence from the fluorescence imaging toolkit. There is great demand in the cell biology community for a red $Ca^{2+}$ biosensor that could be imaged simultaneously with a green fluorescent marker. For example, in the field of receptor biology there is a

widely acknowledged need to monitor receptor localization (with a GFP-fusion to a G-protein coupled receptor) and cell activation simultaneously by monitoring transient $Ca^{2+}$ elevations.

CaM/M13 was inserted into mCherry and the insertion had deleterious effects on mCherry folding and fluorescence. Even after $Ca^{2+}$ was added, the Camgaroos were still non-fluorescent. Two possible approaches to improve them were considered: one is deleting the M13 by PCR, which is still in progress; and the other is random mutagenesis by error-prone PCR. Of course, the two approaches could be combined and used together: deleting the M13 first and then random mutagenesis.

For future work, the Pericam-type variants of mCherry will be constructed by fusion of the $Ca^{2+}$-binding protein calmodulin (CaM) and the CaM/$Ca^{2+}$-binding peptide M13 to the new N- and C- termini of each cpmCherry variant. Such a design proved very effective for construction of green G-CaMP and yellow Pericam $Ca^{2+}$-biosensors. For each cpmCherry variant, the two alternative arrangements of CaM/M13 fusions will be created ($N_{term}$-CaM-cpmCherry-M13-$C_{term}$ and $N_{term}$-M13-cpmCherry-CaM-$C_{term}$) and these variants will be tested *in vitro* for change in fluorescence intensity in the presence or absence of $Ca^{2+}$. If promising variants are identified, the fluorescent response of the $Ca^{2+}$ biosensor will be maximized through the fine-tuning of the linkers between the domains of the biosensor. If promising variants are not identified, 12 variants (6 × cpmCherry variants × 2 arrangements of CaM/M13) will be remade with shorter linkers that are more likely to communicate the structural change to the FP, but may also be

detrimental to folding. If this approach ultimately proves unsuccessful, we will explore transposon-based approaches for creating large libraries of cp-variants containing CaM and M13 fused to the new termini. We intend to demonstrate the utility of our red fluorescent $Ca^{2+}$ biosensor by targeting it to the endoplasmic reticulum (ER) of HeLa cells and performing proof-of-concept experiments.

Furthermore, insertion or fusion of other proteins with molecular-recognition domains, like the zinc finger domain, into mCherry or cpmCherry would be an alternative approach to make other molecular biosensors.

In conclusion, the tolerance of mCherry to circular permutations and insertions shows that the folding process is surprisingly robust and offers a new strategy for creating genetically encoded indicators for important biochemical and physiological signals by fusion of other molecules with molecular recognition domains.

# APPENDIX A:

# TUBES AND LABELS

**Table A.1** cpmCherry variants gene and the gene would be used to make mCherry $Ca^{2+}$ sensors.

| Lable on tubes | Description | |
|---|---|---|
| 1-136 | cpmCherry gene, PCR (mCherry as template) with primers | YL001 and YL002 |
| 9-137 | | YL009 and YL010 |
| 5--21 | | YL005 and YL006 |
| 7--23 | | YL007 and YL008 |
| 3-192 | | YL003 and YL004 |
| 11-183 | | YL011 and YL012 |
| 136-0 | cpmCherry variants (plamid in pBAD) without any mutation | |
| 137-0 | | |
| 21-0 | | |
| 23-0 | | |
| 192-0 | | |
| 183-0 | | |
| cpmCherry2 | | |
| 192-1-0 | 192cpmCherry1.2 (A57V, E39Q) | |
| 183-mix-3 | the brightest one from183-0 mutants named 183cpmCherry1.5(N195D, N6bl) | |
| CaM/M13 | XbaI-CaM-M13-EcoRI (PCR:Primers YL013 and YL014, template: YC3.3 | |
| CaM/M13 vec. | ...pBAD-CaM-Xho1-Kpn1-Nhe1-pBAD.., might be wrong because there is one NheI in pBAD | |
| 183 Xho1/sph1 | Xho1-cp183mix3-sph1 | |
| 183 Xho1/Nhe1 | Xho1-cp183mix3-Nhe1 | |
| pBAD no sph1 | ...pBAD-Xho1-FP-Sph1-CaM-M13-SacI-FP-EcoRI-pBAD.. (from Huiwang) | |
| pBAD no Nhe1 | quick change NheI in pBAD, containing CFP (from Zihao) | |

**Table A.2** mCherry insertion variants gene and the gene was used to make the insertion variants.

| Lable on tube | Description |
|---|---|
| O1 | miniprep of 15bp insertion mCherry Variants, 01 at 138-139, 02 at 137-138, 19 at 25-26, 32 at 22-23, 47 at 193-194, and 67 at 183-184 |
| O2 | |
| 19 | |
| 32 | |
| 47 | |
| 67 | |
| 01 vec | vectors are used to prepare Camgroo. The miniprep (01,02,19,32,47,67, above) were digested with Not1 and modified with CIAP. 47vec and 67 vec have bad backgroup. |
| 02 vec | |
| 19 vec | |
| 32 vec | |
| 47 vec | |
| 67 vec | |
| C01 | CaM/M13 insertion variants. C01,C02, C19 and C32 were conformed by sequence. C47 and C67 were not. |
| C02 | |
| C19 | |
| C32 | |
| C47 | |
| C67 | |
| CaM/M13 01,32,67,19 | CaM/M13 gene used inserted into the vectors above. |
| CaM/M13 47 | |
| CaM/M13 02 | |
| C01-D | CaM insertin at 01(138-139), not confirmed by sequence |
| M1-Kan$^R$ | Components of MGS kits, MuA in Enzyme box. (Lable in red are newer and in black are older.) |
| M1-CaM$^R$ | |
| 5 X buffer | |
| Control DNA | |
| MGS mRFP | M1-Kan (Entranceposon) inserted within mCherry gene |
| MGS+ pBAD CD | the above gene ligated into pBAD |
| MGS + NotI | the above plasmid digested with NotI |
| 15bp I-L | the above gene self-ligation and make 15 bp insertion library from ~10000 colonies |
| circuliarization | circularization mCherry gene, no purified |
| BE library | Blunt end ligation library |
| mRRP1.5 | miniprep |
| mRRP2.0 | mCherry miniprep |
| puc18 | puc18 containing CFP |
| YC3.3 | miniprep |
| puc18 vector | digested with XbaI and EcoRI |
| pBAD EF | miniprep,making by PCR with primers YLEI and YLF1, contains NcoI, NotI and Pst1restriction sites. It is used to prepare blunt end vector. |
| pBAD CD | miniprep,making by PCR with primers YLCI and YLD1, contains XbaI, Kpn1 and EcoRI restriction sites. It is used to prepare blunt end vector. |
| mRFP in puc18 | mCherry Maxiprep |
| mRFP in pBAD | mCherry miniprep, XbaI-Kpn1-GGS-mCherry-G-Kpn1(GT)-EcoRI |

| | |
|---|---|
| mCherry2 in pBAD | mcherry miniprep, XbaI-Kpn1-GGS-mCherry2-G-Kpn1(GT)-EcoRI |
| circuliar mCherry2 | circularization mCherry gene, no purified |
| pBAD XbaI/EcoRI | pBAD CD digested with XbaI and EcoRI |
| **Lable on tube** | **Description** |
| pBAD KpnI/EcoRI | pBAD CD digested with KpnI and EcoRI |
| mCherry2 + gene | mCherry gene made by PCR with primers YLA1 and YLB1 |
| mCherry2 Kpn1 | mCherry2 in pBAD digested with Kpn1 |
| Others | 15 bp insertions good in digest test but sequence is not good enough to tell |

**Table A.3** Primers were used to in this thesis working.

| Sequence | Description | |
|---|---|---|
| ATGCCATAGCATTTTTATCC | pBAD forward sequencing primer(48.4) | |
| GATTTAATCTGTATCAGG | pBAD reverse sequencing primer(41.3) | |
| TGTAAAACGACGGCCAGT | pUC forward sequencing primer(54.4) | |
| AACAGCTATGACCATG | pUC reverse sequencing primer(45.4) | |
| | | |
| ggtctagaggtaccggcggctcc**atg gtgagcaagggcgag** | Xba1_Kpn1_GGS_mchry_fw(58) | xba1-kpn1-GGS-mRFP-G-kpn1-EcoR1 Circular:GTGGS mchry GGT Ramdomdigest:mchry(half) GGTGGS mchry(half) |
| **ggcatggacgagctgtacaag**ggcgg taccgaattcc | EcoR1_Kpn1_GGT_mchry_rv(58) | |
| GGAATTCGGTACCGCCCTTGTACAGC TCGTCCATGCC | | |
| | | |
| ggggtaccgaattctaa**gcttggctg ttttggcgg** | Kpn1_EcoR1_stop_pBAD(fw59.7) | pBAD-xba1-kpn1-EcoR1-stop-pBAD(keephistag) |
| **gacgatgacgataaggatccgagc**tc tagaggtacccc | pBAD_Xba1_Kpn1(rv58.9C) | |
| GGGGTACCTCTAGAGCTCGGATCCTT ATCGTCATCGTC | | |
| | | |
| AACTGCAGGCTAAGTAAGTAA**CTGTT TTGGCGGATGAGAGAAGATTTTC** | pst1-stop-pBAD(fw59.5) | pBAD-Nco1-Not1-pst1-stop-pBAD The PCR product is digested with pst1 and Nco1 and then modified with T4 polymerase to make blunt end vector.(without histag) |
| **GGGCTAACAGGAGGAATTAACCATGG** CGGCCGCTGCAGAA | pBAD-Nco1-Not1-pst1(rw59.2) | |
| TTCTGCAGCGGCCGCCATGGTTAATT CCTCCTGTTAGCCC | | |

| Sequence | Description | |
|---|---|---|
| GGCGGTACCGGCGGCTCC | fw,with RECB1 to test(67.2) | cp mchery linker primer |
| GGAGCCGCCGGTACCGCC | rv,with RECA1 to test(67.2) | |
| TGCGGCCGCA | for MGS mchery,(50.5) | |
| | | |
| ATGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTG | pBADfw(63.4),using with YLH1 | Because the Tm of RECB1 and RECA1 is too low. Use these two place them for PCR |
| CCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTG | pBADrv(63.4),using with YLG1 | Colone CaM+M13 fromYC3.3 The frame for 01,32,67,19 |
| AAGGAAAAAA**GCGGCCGCAGACCAACTGACAGAAGAGCAGATTG** | fw(57.6) | |
| AAGGAAAAAA**GCGGCCGCCAGTGCCCCGGAGCT** | rv(58.2) | The frame for 02 |
| AAGGAAAAAA**GCGGCCGCTAGACCAACTGACAGAAGAGCAGATTG** | fw(57.6) | |
| AAGGAAAAAA**GCGGCCGCCCCAGTGCCCCGGAGCT** | rv(58.2) | The frame for 47 |
| AAGGAAAAAA**GCGGCCGCGACCAACTGACAGAAGAGCAGATTG** | fw(57.6) | |
| AAGGAAAAAA**GCGGCCGCGCAGTGCCCCGGAGCT** | rv(58.2) | Deleting M13 |
| AT**GGCGCCCTTTGCTGTCATCATTTGTACAAACTCTTCATA** | 57.6CaMrvCaM_Nar1 | |
| AT**GGCGCCAAGACCATGGGCTGGGAG** | 57.2rw | |
| AT**GGCGCCAACGGCCACGAGTTCG** | 55.6rw | |
| AT**GGCGCCAACGTCAACATCAAGTTGGACATCAC** | 57.8rw | |
| AT**GGCGCCAAGCCCGTGCAGCTG** | 56.3rw | |
| AT**GGCGCCTTCGAGATCGAGGGCGAGG** | 59rw | |

| Sequence | Description | |
|---|---|---|
| CCGAACTCGAG**GACCAACTGACAGAA GAGCAGATTG** | 57.6CaMrvCaM_Xho1 | Making pericam mchey from the insertion variants.Butthere is some thing wrong,so change to YLY1andYLZ1 |
| GGAATTC**AAGAGGGAGTGGGAGAAAA ACTTCAT** | 58M13fwEcoR1_M13 | I haven't ordered them,yet. |
| CCGAACTCGAG**AAGAGGGAGTGGGAG AAAAACTTCAT** | fw,Xho1_M13 | |
| GGAATTC**CTTTGCTGTCATCATTTGT ACAAACTCTTCATA** | rv,CaM_EcoR1 | |
| | | |
| **Making Cicularlly Pemuted Mcherry** | | |
| CCGAGTCTAGA**ATGCAGAAGAAGACCATGGGC** | | fw,Xba1-13 |
| TTCGAATTC**TTACTTCTGCATTACGG GGCCGTC** | rv,138-STOP-EcoR1 | still in solid, use to make cpmecherry at 215-216, which was serched by blunt end ligation library. |
| | | |
| CCGAGTCTAGA**GCCTACAACGTCAAC ATCAAG** | fw,Xba1-192 | |
| TTCGAATTC**TTAGTTGTAGGCGCCGG GCAG** | rv,194-STOP-EcoR1 | |
| | | |
| CCGAGTCTAGA**TCCGTGAACGGCCAC GAG** | fw,Xba1-21 | |
| TTCGAATTC**TTAGTTCACGGAGCCCT CCATGTG** | rv,23-STOP-EcoR1 | |

| | | |
|---|---|---|
| | | |
| CCGAGTCTAGA**CACGAGTTCGAGATC GAGGGC** | fw,Xba1-25 | |
| TTCGAATTC**TTAGAACTCGTGGCCGT TCACGG** | rv,27-STOP-EcoR1 | |
| | | |
| **Sequence** | **Description** | |
| CCGAGTCTAGA**CAGAAGAAGACCATG GGCTGG** | fw,Xba1-137 | |
| TTCGAATTC**TTACTTCTTCTGCATTA CGGGGCC** | rv,139-STOP-EcoR1 | |
| CCGAGTCTAGA**GCCAAGAAGCCCGTG CAG** | fw,Xba1-183 | |
| TTCGAATTC**TTACTTCTTGGCCTTGT AGGTGGT** | rv,185-STOP-EcoR1 | |
| CCGAGTCTAGA**CGCGCCGAGGGC** | fw,Xba1-215 | |
| TTCGAATTC**TTCGTACTGTTCCACGA TGGTGTAG** | rv,216-EcoR1 | |
| | | |
| **Making Ca Sensor** | | |
| ccgagTCTAGA**CAGCTCCGGGGCACT G** | fw,Xba1-CaM | |
| TTCGAATTC**CAGTGCCCCGGAGCTG** | | rv,m13-Eco |
| | | |
| ccgagctcgag**GCCAAGAAGCCCGTG CAG** | xho1-183 | |
| TTCGCTAGC**CTTCTTGGCCTTGTAGG TGGTCTTG** | 185-Nhe1 | Xho1-cpmchery-Nhe1 |
| | | |
| Ccgagctcgag**gcctacaacgtcaac atcaagttgg** | xho1-192 | ..pBAD-CaM-Xho1-Kpn1-Nhe1-pBAD... Note: there is Nhe1 site in pBAD, |

| | | |
|---|---|---|
| TTCGCTAGC**GTTGTAGGCGCCGGGCA G** | 194-Nhe1 | Changed to use sph1 |
| ccggtaccgctagc**aagagggagtgg gagaaaaacttcattg** | Fw,Kpn1-Nhe1-M1358.8 | |
| ggggtaccCTCGAG**GCTGTCATCATT TGTACAAACTCTTCATAGTTTACTTG** | rv,CaM-Xho1-Kpn159.1 | |
| **Sequence** | **Description** | |
| TTC**GCATGC**CTTCTTGGCCTTGTAGG TGGTCTTG | 183-sph1 | |
| TTC**GCATGC**GTTGTAGGCGCCGGGCA G | 192-sph1 | |
| CCGGTACC**GCATGC**AAGAGGGAGTGG GAGAAAAACTTCATTG | Kpn1-sph1-M13 | |
| GGGGTACCCTCGAG**CTTTGCTGTCAT CATTTGTACAAACTCTTCATAGTTTA** | rv,CaM-Xho1-Kpn1 | |
| | | instead of YL020 |
| | | |
| | | |